AI & SOCIETY (2019) 34:355–364
https://doi.org/10.1007/s00146-018-0854-4

OPEN FORUM

Posthuman learning: AI from novice to expert?

Cathrine Hasse[1]

Received: 15 June 2017 / Accepted: 10 June 2018 / Published online: 3 July 2018

Abstract
Will robots ever be able to learn like humans? To answer that question, one first needs to ask: what is learning? Hubert and Stuart Dreyfus had a point when they claimed that computers and robots would never be able to learn like humans because human learning, after an initial phase of rule-based learning, is uncertain, context sensitive and intuitive (Dreyfus and Dreyfus in A five stage model of the mental activities involved in directed skill acquisition. (Supported by the U.S. Air Force, Office of Scientific Research (AFSC) under contract F49620-C-0063 with the University of California) Berkeley, February 1980. (Unpublished study). Washington, DC: Storming Media. https://www.stormingmedia.us/15/1554/A155480.html. Accessed 10 Oct 2017, 1980). I would add that learning also builds on prior learning, and that from the outset (birth), human learning is a socio-cultural materially grounded collective epistemology. This posthuman acknowledgement shifts the focus from the individual learner to learning within collective phenomena. Dreyfus and Dreyfus (1980) do not seem to emphasise the essentially social and cultural nature of the human condition. Learning theory (especially the Vygotskyan perspective), new materialism (especially as presented by the physicist Karen Barad) and postphenomenology (especially as presented by Don Ihde) have emphasised in different ways the materially based socio-cultural nature of human learning. They thereby point towards a 'posthuman' learning that is far from the machine-like or enhanced creature envisioned by singularists. Until robots are essentially social and ground their epistemologies in socio-cultural materiality, I suggest that human-like AI is not possible.

Keywords Learning · AI · Dreyfus · Stage model · Posthumanism

1 Introduction: AI from novice to expert?

In 1980, the philosopher Hubert L. Dreyfus, together with his brother, the mathematician Stuart Dreyfus, wrote a research paper on how, with adequate instruction, airline pilots, language learners and chess players could move from novice to expert (Dreyfus and Dreyfus 1980). The paper was entitled *A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition*. Though the stage model has later become renowned as a theory of learning, it was at first rather a modelling of the outcome of learning new skills. In several later versions, the two brothers, individually and together, expanded their original model with a number of additions, of which I shall take a closer look at two: how skills and expertise build on learning and how this process of learning is particularly human, of a kind that so far cannot be replicated by machines. New developments in learning theories that connect socio-cultural theories of learning with new materialism and postphenomenology add yet another dimension to this argument that I shall term *posthuman learning*. The concept of the posthuman has become popular in recent years, but confusingly, it is used in two ways. The first way is connected to the 'singularist' movement, named after Ray Kurzweil's arguments in the book *The Singularity Is Near: When Humans Transcend Biology*, and it claims of an imminent future where not only can machines learn, they become better learners than humans, resulting in the rise of a new species of machine-bodied posthumans that surpass humans (Kurzweil 2005). The other definition of 'posthuman', which I draw upon in this article, refers to a redefinition of 'the human' as emphasised by new materialists such as the feminist physicist Karen Barad (2007) and the feminist philosopher Rosi Braidotti (2013). From this perspective, the human is not a stand-alone individual engaging with a world of discrete objects, as has been the belief since the

✉ Cathrine Hasse
caha@dpu.dk

[1] Danish School of Education, University of Aarhus, Tuborgvej 164, 2400 Copenhagen, NV, Denmark

enlightenment, but a posthuman 'coming-into-being' with socio-cultural materiality. This point has also been made by socio-cultural learning theorists following the work of the Russian psychologist Lev Vygotsky who emphasised that what we have already learned culturally (prior learning) is a condition for new learning. This approach to learning, on the other hand, lacks the constantly present sensing human body emphasised by the Dreyfus brothers and postphenomenologists such as Ihde (e.g. 2002).

In the following paragraphs, I first present the Dreyfus brothers' model and review the 'human' in their original arguments. Next, I introduce the new materialist concept of 'the posthuman' and connect this new understanding of humans with Vygotsky's socio-cultural learning theory. Reconsidering the Dreyfus brothers' arguments in light of these perspectives opens up for a new understanding of the learning process in the move from novice to expert. Neither the novice nor the expert is a human learner in the traditional understanding of individual human learners, but they become learners in a process that connects bodies, word meaning and materiality in an ongoing cultural learning process. As an inherently socio-cultural material epistemology, this theory of learning can be used to argue that machine learning in AI builds on an outdated paradigm of the detached, rational human. Posthuman learning supplements the Dreyfus brothers' arguments with an understanding that both the novice and the expert become skilful in an evolving process of collective socio-cultural material epistemology.

2 The five-stage model

When the Dreyfus brothers presented their first version of the stage model in 1980, they built it on research on how airline pilots learned to fly planes and several earlier drafts developing the thoughts behind the model (Dreyfus and Dreyfus 1980, 6). The five stages they identified at first progressed from novice, competent, proficient, expert to master and they saw these stages as connected to differences in the following mental functions: recollection, recognition, decision and awareness. Although the concept of learning was mentioned in this first version, it was invoked somewhat instrumentally and not treated as a concept in need of scrutiny. Furthermore, it was not explained how learning was connected to recollection, recognition, decision and awareness. Always wary of the pitfalls of understanding mental activity as information processing, the Dreyfus brothers suggested that proficiency moved from abstract to concrete rather than in the opposite direction:

> Rather than adopting the currently accepted Piagetian view that proficiency increases as one moves from the concrete to the abstract, we argue that skill in its minimal form is produced by following abstract formal rules, but that only experience with concrete cases can account for higher levels of performance (Dreyfus and Dreyfus 1980, 5).

In later expansions of the model, they argued that in the novice stage the rules and elements to be learned were so explicitly and clearly presented that they were "context-free" (Dreyfus and Dreyfus 1986, 21). The novice stage is characterised by clear human–human instruction, where the instructor introduces these "context-free" rule-based features. Furthermore, in 1986 the brothers again emphasised that the novice stage was close to rule-based machine learning. In his 2004 summary of the stage model, Stuart Dreyfus even writes that the novice "just like a computer following a program" learns the "context-free" features (Dreyfus 2004, 177). Where the first text from 1980 focussed on airplane pilots, chess players and second language learners, the brothers also introduced car drivers in their 1986 version. I will stick to this example in my presentation of the model—and even though the brothers continued to work on it—I will here present the model in the version found in the brothers' work from 1986 and in the summary by Stuart Dreyfus in 2004.

Stage 1 Novice: the novice learns the 'context-free' features of car driving, such as that speed is indicated by the car's speedometer. She learns rules, for instance concerning the meaning of signs along the roadside or that she should shift gears when the needle of the speedometer reaches a certain number according to speed limit regulations. This is 'context-free' information in so far the novice has yet to learn what it is like to drive a car in practice.

Stage 2 Advanced beginner: as she begins to drive the car, the "context-free" information is exchanged with situated insights. She no longer blindly follows the speedometer, but includes the sounds of the engine in her judgement of speed. She has become an "advanced beginner" (Dreyfus and Dreyfus 1986, 23).[1]

Stage 3 Competence: in the next phase of learning, many inputs from "context-free" information and situated environment become overwhelming, and it becomes necessary to learn how to discriminate between relevant and irrelevant information. Instruction and experience over time help build up a perspective that makes it possible to choose the relevant and disregard the irrelevant features in the surroundings. The competent car driver will still pay attention to rules, but now adopt them to fulfil a plan and to achieve a goal. The driver begins to make decisions in the situation that result

[1] Note, I have reversed the gender presented by the Dreyfus brothers, who themselves lamented that they were 'painfully aware' of their use of 'he' (Dreyfus and Dreyfus 1986, 20).

less from the driving in itself, but are tied to why the driver is driving (decisions can be made because the driver needs to go somewhere in a hurry, making her skip some of the precautions prescribed by the rules). She enters into many new situations and considers many options not covered by manuals. Thus, she learns that the situated practice cannot be plotted into a formula and that she will have to live with a degree of uncertainty.

Stage 4 Proficiency: at this stage, the learner experience what the model calls 'holistic similarity recognition' (Dreyfus and Dreyfus 1986, 28). As the competence turns into proficiency, the driver become more and more intuitive in her driving, and rarely stops to remember rules or even plans in relation to how to drive to obtain a goal. Her actual driving becomes embodied (Dreyfus 2004, 179). The proficient driver stops reflecting on emotionally laden, problematic situations of the past to make choices. The proficient driver simply feels that the car is going too fast around a steep curve. Instead of the competent driver's attention to the relation between separate elements such as the speed, time, angles of road banks and the changing gravitational field, she can almost immediately recognise and decide between a number of possibilities for dealing with the situation (whereas the competent driver used these elements to decide that she was speeding) (Dreyfus 2004, 179).

Stage 5 Expertise: the final stage fulfils the process of embodiment. From the different more or less detached and reflected observations, driving become so intuitive that the driving is not even noticed by the driver. Decisions are not made consciously. Where the proficient performer deconstructs situations and thus allows an immediate intuitive response to each recognised situation, the expert driver, generally without any explicit attention to the surroundings, "not only knows by feel and familiarity when an action such as slowing is required, but generally knows how to perform the act without evaluating and comparing alternatives" (Dreyfus and Dreyfus 1986, 32–33). "What must be done, simply is done" (Dreyfus 2004, 180).

3 The human in the model

The human in the stage model is not a rational, symbol-processing machine-like creature. It is not rational in the sense defined by John McDowell as: "rationality as detached, brought to bear on practical predicaments from a standpoint other than one of immersion in them" (McDowell 2007, 338). [2] It is a phenomenal sensing human; that is, a being-in-the-world. It is a human that initially learns rule-based, like a computer, but soon turns into an embodied learner, whereby learning increasingly stems from situated experiences.

Inspired by the philosophy of perception presented by Maurice Merleau-Ponty and the philosophy of being presented by Martin Heidegger, from his early days as a philosopher at MIT, Hubert Dreyfus was fascinated by human knowledge and perception as emerging from body–world relations. He teamed up with his brother Stuart to do research for the research foundation RAND and it caused quite a stir when he, as early as 1965, published a report that openly attacked a number of other researchers also funded by the RAND foundation. The researchers, among them Marvin Minsky, Herbert Simon and Allen Newell, were hired to do research in AI—artificial intelligence. Dreyfus used their work on artificial intelligence to nail down what a human was not: a human was not learning, knowing and perceiving like a machine. The machines at the MIT laboratories for artificial intelligence operated according to symbolic rules find solutions to purely formal tasks—whereas human intelligence, Dreyfus argued—is embodied and situated (Dreyfus 1965).

In *Alchemy and Artificial Intelligence*, Dreyfus argued that AI was like modern alchemy for two reasons, as summarised by Margaret Boden:

> On one hand, it declared the overall project to be in principle impossible for philosophical reasons, in essence, because the ''higher'' forms of intelligence are necessarily derived from ''lower'' forms concerned with bodily action (…) On the other hand, it mocked the performance of the programmes that had actually appeared thus far. Dreyfus accused NewFAI of four general performance failings, each missing out some ''essential'' aspect of human intelligence. These were reliance on the fringe of consciousness, discrimination between the essential and the accidental, tolerance of ambiguity, and perspicuous grouping (Boden 2006, 839).

The human in the Dreyfus model, unlike machines, can exist in a vague world such as the world described by phenomenology as well as postphenomenology. In this world, background sounds, such as those of a car engine, form part of our existence at the 'fringe' of our consciousness. Humans can discriminate between a rising and falling hum of a passing factory while driving and a car engine that slowly revs

[2] The Dreyfus–McDowell debate (see McDowell 2007) is more complex than I can deal with here, even if relevant for the topic of learning. It is about if and how rationality is already embodied in perception. The cultural aspect of materiality is often overlooked in these debates and that the 'affordances' McDowell and Dreyfus argue about could be seen as culturally diverse intra-active meetings halfway between concept formation and unpredictable materials.

up and down. In such ways, humans differ from machines in Dreyfus's 1965 paper in terms of their abilities to:

> (1) Distinguish the essential from the inessential features of a particular instance of a pattern; (2) use cues which remain on the fringes of consciousness; (3) take account of the context; (4) perceive the individual as typical, i.e. situate the individual with respect to a paradigm case (Dreyfus 1965, 45–46).

Thus, Dreyfus's 'human' emerges as an ambivalent creature, partly rule-formed novice, partly intuitive—but first and foremost a being for which "the body plays a crucial role in making possible intelligent behavior" (Dreyfus 1965, 59).

However, the human subject presented by Dreyfus throughout his work has itself been criticised by, among others, Maxine Sheets-Johnstone. She posed the question whether Dreyfus, despite emphasising a phenomenological approach, also has a somewhat Cartesian understanding of human learning as taking place in neural networks and brain dynamics, rather than being a truly embodied process (Sheets-Johnstone 2000, 357). Contrary to computers, humans do not need someone (a programmer) to distinguish the essential from the accidental. Human pattern recognition, Dreyfus argues, does not rely on the application of fixed rules and such rules are, therefore, not necessary for intelligent human behaviour. Humans need not process internal representations to be intelligent. Intelligence is context-specific and embedded in local situated practices. Even so, as noted by Sheets-Johnstone, the Dreyfus brothers' 'human' is very much a mental being. Although the body is emphasised in their work, the AI community the Dreyfus brothers argue against, also seems to have inspired the 'human' they present.

In Hubert and Stuart Dreyfus' showdowns with AI researchers, and especially in Hubert's arguments, the human–world relation is not as envisioned by the engineers. Where engineers see humans as engaging with a separate material world that exists completely independent of such human engagements with it, the phenomenological approach proposed by Dreyfus emphasises how mental activity changes, for example, for car drivers during skill acquisition. Humans are not beings attached to a pre-existing world, rather, they and the world become together in a process where the world and the human constitute each other. However, the process discussed in the stage model is and remains an individual process—both gears, instructors, manuals and sharp corners are seen from a first-person perspective—and thus learning is an individual process where the emphasis is on how mental skills transform the individual human's relation to a surrounding material and social world.

In spite of their fierce attack on the AI community, there are also some overlaps with the Dreyfus brothers' perception of learning as belonging to mental, first-person processes. To develop some of Hubert Dreyfus's critique of AI, I will now take a closer look at more recent posthumanist theories and what they can tell us about AI and machine learning—as well as sketch a new approach to learning as posthuman.

4 The posthuman

The concept of 'posthuman' emerged in the wake of the 'post' era from the 1980s and onwards, accompanied by a number of other posts, such as post-modernism, post-colonialism, post-feminism and postphenomenology. Like all the other posts, posthumanism both departs from and builds on that of which it is a post. As defined by the philosopher Francesca Ferrando:

> In contemporary academic debate, "posthuman" has become a key term to cope with an urgency for the integral redefinition of the notion of the human, following the onto-epistemological as well as scientific and bio-technological developments of the twentieth and twenty-first centuries. (Ferrando 2013, 26).

From a learning perspective, it is decisive which human we are post (Ihde 2011). Is it the rational, detached human that will be biologically and technically enhanced, as argued by the singularists (Kurzweil 2005) and transhumanists (More and Vita-More 2013)? Is it the rational human of the enlightenment with his [deliberate gender] understanding of himself as the master of the universe with a monopoly on intelligence, rationality and a right to exploit all minor creatures and things on Earth? This is the human Dreyfus argued against—a human whose existence could be reduced to an information computing intelligence that can be surpassed by machine intelligence that is better at computational thinking. This position is also shared by phenomenologists such as Sheets-Johnstone, as well as postphenomenologists like Don Ihde. They all emphasise an embodied subject that stands in a world relation with their kinetic tactile-kinaesthetic bodies. However, these bodies are, so to speak, already in place as individual bodies in an 'intersubjective' world. I add to this that, from a learning perspective, we are also collective learners embodied in collective surroundings. These surroundings are not just conceptual, but also material. When we learn, we not only transform our mental processes or bodies, but the material we engage with as well—just as materials transform us. These materials can be living or non-living. Even more radically, following the female physicist Karen Barad, living and non-living subjects and objects do not pre-exist each other. If we follow her arguments, Dreyfus did not go far enough in his critique. He did not grant the material surroundings their due in embodied skills creation, and furthermore, from my posthuman learning perspective, he saw skills from an individual first-person perspective.

Following Kathrine Hayles (1999), the posthuman could simply be a new conception of the humans we always were. What is 'post' is how we perceive these humans—no longer as individual, stand-alone, rational and privileged creatures of a free-floating information-based intelligence, but as entangled in and responsible for the world. However, in some posthuman theories there is more to the posthuman than a mere change of concepts. In her book *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning* (2007), Barad builds on insights from other feminists, such as Judith Butler and Donna Haraway, both of whom have argued that the formation of the subject is a performative formation of material bodies. However, she takes a (huge) step further in combining these insights with theories drawn from physics and especially the physicist Niels Bohr. What springs out from these unusual combinations is a new theory of the posthuman as "the ontological inseparability of intra-acting agencies" (2007, 206). Like postphenomenologists such as Don Ihde and Peter-Paul Verbeek, she thus proposes to rethink the subject. However, whereas Ihde suggested replacing the subject with 'embodiment' (2003), a position not so far from the one taken by Dreyfus, referring to our inter-relation with the world, Barad's relational ontology is not an interactive, but an intra-active relation.

5 Intra-active driving

Let us consider the driver going from novice to expert from this new angle. What has changed? According to Barad, the whole process of movement from novice to expert has to be rethought. Matter, like the gear, the speedometer, etc., is no longer a mere instrument for driving. It is, just like the driver herself, constituted in the performance of driving. Following Bohr's arguments, matter is not an inert substance that can be exploited and used because it adheres to predictable causal forces. In the relational ontology, matter is also unpredictable, volatile and vital. The gear may suddenly break, the road may stop, and the tyre may explode. The driver herself is no longer a separate individual human with a privileged position, but is already enmeshed in a cultural material world that makes it possible for gears and tyres to come into being as phenomena. Similarly, technology is not a separate entity that pre-exists these engagements. Furthermore, our entanglements are so culturally diverse that we may speculate that posthumans are not to be regarded as belonging to a single species. This approach makes Barad's posthumanism radically different from the posthumanism proposed by AI-inspired communities (e.g. Ray Kurzweil 2005; Max and Vita Moore 2013).

Ferrando explains it as follows:

> For instance, in the case of chattel slavery, slaves were treated as personal property of an owner, to be bought and sold. And still, transhumanist reflections, in their 'ultra-humanistic' endeavours, do not fully engage with a critical and historical account of the human, which is often presented in a generic and 'fit-for-all' way. (Ferrando 2013, 28).

The driver as a human subject can be gendered (as noted by the Dreyfus brothers themselves), as can a host of other beings for which we have words like race, age, criminal, lawyer, etc. In Barad's relational ontology, *there are no fixed representations of material bodies* (humans and non-humans), but:

> specific (re)configurings of the world through which boundaries, properties, and meanings are differentially enacted (i.e. discursive practices, in my posthumanist sense) and specific material phenomena (i.e. differentiating patterns of mattering). (Barad 2007, 139).

The dynamism of matter is not due to an inherent property of things (brute nature), but a process with shifting entangled relations emerging from within phenomena rather than within an individual. Dreyfus noted that driving is comprised of a series of phenomena which result in a process of knowing. However, knowing is not reducible to a mental process; knowing is rather a *physical practice of engagement* (Barad 2007, 342).

The driver and the surroundings can, of course, be analytically separated, and as such, Barad would grant them existence as subject–object differentiation within phenomena. Though Barad herself claims (in a note, Barad 2007, n. 30, p 412) that connotations to phenomenology are 'unwanted', there is a lot of phenomenological thinking to be found in her work, especially in relation to Merleau-Ponty. However, her radical insistence on matter's dynamism creates a new agential account of shifting power relations. The phenomena that comprise driving may include the driver's body intra-acting with the gear to become a worn-down ready to dismiss gear case, or the gear and the driver intra-acting with the driver's view of herself as a bad gear-shifter. Whatever is discriminated is within intra-action. What is included in these entangled intra-actions will always be a matter of subsequent analysis, which for Barad is a new entanglement. It could be the slippery surface of the road, but it could also be the contractor who was supposed to repair the road, but did not because it was raining that day, or it could be a child who left a spike just to see car tyres exploding, or a piece of plastic landing on the screen, blown by a wind caused by global warming. The driving phenomena may not be confined to the instructor, the car and the driver—and may be even more situated in the contingencies of everyday life than as revealed in the Dreyfus brothers' arguments.

6 Restoring the learning body

The new feminist materialists like Barad have emphasised the 'intra-active' relation between words and world. They have also dismantled the problematic conception of the 'human' as a stand-alone, rational and intelligent individual (Braidotti 2013). In its place, they have the 'posthuman', which is no longer a fixed entity or subject, but a shifting entanglement of humans and non-humans. Though I am inspired by Barad's radical intra-activity, the 'intra' also poses some problems that are tied to the original argument made by the Dreyfus brothers. These problems concern learning as the basic process that transforms our awareness of a world that is recognised, remembered and where we also (sometimes) make decisions. I acknowledge that matter as a dynamic force has been granted too little attention in phenomenology, and even postphenomenology—and I can accept that matter (including human bodies) is not a demarcated entity with fixed boundaries, but always comes into being as part of a phenomenon. However, even in Barad's radical universe, if concepts are involved at all, they must include an evolving human embodied consciousness (through learning). This process of prior learning is always included in phenomena whenever humans are involved.

In the rest of this article, I will argue that the materialist posthumanists have left out an important aspect of these entanglements, namely, collectively embodied mental processes of prior learning. Like phenomenology and postphenomenology, Barad's relational ontology is not built on an account of established boundaries, for example, between nature and culture. Yet, the mental is not the same as the material brain (see e.g. Nath and Sahu 2017). Human psychological processes are the pivotal points for creating phenomena and it should not be ignored how values, languages and practices are tied to cultures (Gill 1991).

Though brains definitely matter—as when a driver loses the ability to drive after developing Alzheimer's—the mental cannot be reduced to the brain. The mental is not 'subjective'—as our individual perceptions of the colour 'red'. From the posthuman learning perspective that I propose (see also Hasse 2015), the posthuman is not an individual 'subjective' human, but a collective 'subject'. Drivers can disregard a red light, but they have all collectively learned to recognise it as a sign to stop—if they belong to a driving community. Some humans, even today, live in areas without cars and will not know a 'red' light means stop in the way experienced drivers do. When the novices move towards expertise, their world changes in cultural ways that gradually align their experiences with a whole community of drivers. It is through this cultural learning process we become engaged.

We cannot experience the world as a bat, as argued by Thomas Nagel (1974), but we do learn to experience the world as many other people around us. Experience is thus not just subjective in an individualist way, but to some extent collectively shared through our socio-cultural learning processes that merge words, meanings and materials in ways that make social communication possible. This argument was first made by the Russian learning theorist, Lev Vygotsky, in the 1920s (Vygotsky 1987). However, these learning theories have had little impact in AI environments, where the preferred learning theorist has been Jean Piaget, inspiring both Seymour Papert and Marvin Minsky—and influencing their instrumental conception of learning: *"Learning" is making useful changes in the workings of our minds*" (Minsky 1986, 120). In the academic field of learning theory, there are many debates about the differences between the respective concepts of learning in the work of Piaget and Vygotsky. The most important is that of 'inner speech'.

For Piaget, the development of a human is tied to a concept of an individual human that is not so far removed from the 'human' emulated by AI. This human is initially a thinking being, with certain structures in place, before she learns to communicate and socialise. For Vygotsky, thinking and speech develop together through social and material interactions (1987). Although this human is also an individual, the individual subject is first formed by the surrounding collectives—sociality in the shape of other humans (like driving instructors) and the social consciousness embedded in material surroundings (like cars, roads, and signs). This human is of a different kind than the rational one objected to by the posthumanists and the Dreyfus brothers. While Vygotsky himself did not go that far, based on his thinking, we can outline a cultural learning theory in which we cannot take what humans share for granted, but can explain the diversities (in material worlds and how they are perceived by persons) with reference to learning through different embodied engagements (Hasse 2008).

Vygotsky's human is already a 'posthuman' in so far as she is a collective before she is an individual. When she makes apparently individual decisions on how to walk or drive, her cultural learning processes have already taught her to place her body on the pavement and the car in the right (or left) lane. Furthermore, her mental prior learning makes her perceive the road, gears, steering wheel and the clock at the dashboard as what Vygotsky terms 'real', i.e. 'meaningful', objects. When she perceives the speedometer or the clock in the car, she does not just see a round white circle with black dots, but an object with a collectively shared meaning.

> A special feature of human perception—which arises at a very young age—is the perception of real objects. This is something for which there is no analogy in animal perception. By this term I mean that I do not

> see the world simply in colour and shape but also as a world with sense and meaning. I do not merely see something round and black with two hands; I see a clock and I can distinguish one hand from the other (Vygotsky 1978, 33). [3]

Where the AI sciences deal with abstractions, the drivers (including the AI researchers themselves) deal with meaningful real objects. For Vygotsky, science is abstract thinking as compared to the everyday thinking that is always connected to perception through word meaning:

When we meet what is called a cow and say: "*This is a cow", we add the act of thinking to the act of perception, bringing the given perception under a general concept* (Vygotsky 1987, 250).

We can even go further than Vygotsky, and his distinction between science as dealing with abstract thinking contra everyday thinking, and with Barad, emphasise that materials are always involved and, with phenomenology and postphenomenology, that bodies are always involved. As also argued by Don Ihde, bodies are embedded in technology (see *Bodies in Technology* 2002)—also in a literal way. Our corporeal bodies are also our *corps vecu* (Ihde 2003). In the posthuman learning perspective, this means that technology is not just a mediator, but takes part in the co-creation of a collectively shared socio-cultural material world. Even seemingly abstract symbols are in this view materials entangled in phenomena that include the prior learning in the body of the driver as well as a whole community of drivers.

Driving is no longer an individual process as the driver is already a collective of material and social meaningfulness. Even when she begins as a novice, her prior learning is just as entangled with her driving as the instructor and the gears. She may not know how to drive, but like other humans in her local community, she knows what driving is.

With his emphasis on learning as cultural, collective word meaning in a material world, Vygotsky goes much deeper than perceiving learning as the 'useful changes in the workings of our minds'. Learning is not just a change of 'mind', but of a somewhat collective bodily becoming of a material world. This acknowledgement makes it possible to sustain Hubert Dreyfus's claim that AI will never succeed in making truly thinking machines (from 1965 onwards).

In his critique of AI, Dreyfus explored and refuted some of the assumptions made by scientists engaged in AI, which I shall now review in light of the 'collective driver'. The assumptions that Dreyfus claimed were behind much AI work were: assumptions about biology, psychology, epistemology and ontology Dreyfus (1979).

In the psychological and biological assumptions, the brain of the computer and the human are both expected to work by some kind of mechanism that opens or closes gates and later eases or loads weights. The mind conflates with a brain that operates like a machine, processing incoming bits of information that are stored in relation to formal rules. Though new research in the field of neurology has confirmed that Dreyfus was right when he claimed the brain was more plastic than a computer, and the AI community has largely accepted this view, most AI—even the most advanced machine learning systems—still operates according to systems of opening and closing.

> Deep down we find nothing, but formalised linear step-wise instructions (algorithms) and data that are represented by strings of binary numbers for necessary calculations to be performed. These rules that govern the self-dictating are realised and inscribed in machines by their designers. Thereby it appears that such machines are automatic instead of autonomous in the sense that they do not possess the right or condition of self-government in the sense as a free person does, namely, constitutive autonomy. (Lyyra 2015, 9).

This is the basic and rule-based so-called 'learning' that drives machine learning. Humans do not learn formalised knowledge in this manner, as assumed by (some) AI scientists and refuted by Dreyfus. Dreyfus likewise refutes the ontological assumption that the world consists of separate symbols and objects—and that the relation between the two can be formalised by symbol processing. From a posthuman learning perspective, we do not learn 'symbols', but as argued by Vygotsky, meaningful words and signs. As also emphasised in classical phenomenology, we are condemned to meaning (Merleau-Ponty 1962).

Not much work has been done to understand the processes (e.g. learning processes) of humans as collective and cultural learners as compared to machine learning, where hundreds of machines (e.g. self-driving cars) can run on the same formal logics. What emerges in intra-actions with a constantly transforming material world (driving today in the rain is different from yesterday's drive in sunny weather) matters to humans in ways that result in different culturally informed responses that would not be meaningful to machines.

With Vygotsky, we get an apparatus to understand how a human-centred approach (e.g. Gill 1991) to learning is unavoidable—only I suggest replacing 'the human' we once were with an emphasis on 'posthumans'. This includes the collective learning of humans in the entanglements suggested by Barad—but the emphasis is on the changing posthuman subjective experience of, e.g. driving. Thus, it is not just the entanglements of material-discursive intra-actions, but also a process that includes how the prior learning of

[3] From a posthuman perspective, it is not at all certain that this is an ability of humans alone. We increasingly acknowledge that animals also live in meaningful worlds.

a 'post' human is already implicated in a material world through collective cultural learning processes.

Learning is a process that can move a person in a car from being a novice to an expert driver. Even when matter behaves in its most dynamic way, the learning process literally matters for the outcome. An experienced or expert driver deals with the flying plastic or an exploding tyre better than the novice does. This brings me back to an argument for what learning is that moves beyond the linear and individual process proposed by the Dreyfus brothers.

First, learning is relational in a cultural sense. Good driving, and even the materialised car itself, is a cultural phenomenon in the sense that not all humans on Earth that have ever heard of, let alone driven, cars accept the same driving as 'good'. If we take the dynamic matter as a point of departure, our mental and material world depends on learning through and with a socio-cultural material world. When available materials shift, so does our potential for learning, and not just learning to drive, but also all the collectively shared thinking, memories and awareness about driving. This collectivity cannot be in the entanglements with materiality without learning, and learning depends on the cultural constellation of available cars, instructors, roads—and a society that values and makes rules for driving. Rules for driving also differ culturally (the English have a different placement of the road than the French, and in Saudi Arabia women have, until recently, not been allowed to drive). This is a culture–culture diversity, rather than the nature–culture diversity that occupies Barad. The instructor is not a neutral person, but like the driver, is formed by dynamic materials and learning. The novice is a person who, on the one hand, has not yet learned to drive, but on the other has learned an enormous amount of knowledge about driving that is already entangled in the situation when she begins to take driving lessons.

Learning is a process that involves culturally informed social engagement that simultaneously transforms who we are, how we perceive and what we are aware of as well as transforming the material world. This is not emphasised enough, neither by Barad, nor in phenomenology and post-phenomenology—doing so will ultimately make it possible to argue, with Dreyfus—that machines do not learn like humans.

7 Conclusion: reviewing the five-stage model

In 1980, the Dreyfus brothers wrote their paper arguing that the highest level of expertise is the intuitive stage, where learning becomes embodied. Their five-stage learning model, leading the novice to the stage of expert, has received wide acclaim, as well as critique, within educational settings. I have reviewed the model from the perspective of a new posthumanist theorising of learning that emphasises learning as an embodied, yet collective, process that entangles humans and non-humans. I suggest that we need an enhanced understanding of 'the human' along the lines proposed by Hubert Dreyfus to grasp the major difference between machine learning and collective human learning. Even if the five-stage model has its flaws in explaining the complexity of human learning, it may still be used to support the critique of AI raised by Hubert Dreyfus. Machine learning has acknowledged that humans do not learn through representational processing of information. However, artificial intelligence has yet to move beyond the enlightenment-informed understanding of the endpoint of human learning as a rational process rather than, as proposed by Dreyfus, an intuitive process. The rational Vitruvian Man is dead, as argued by posthuman proponents (e.g. Braidotti 2013), and so, therefore, are our assumed rights and capabilities to transfer equally assumed intelligence to machines. Furthermore, whenever humans are involved in phenomena, we need to acknowledge that we are dependent on our phenomenal bodies. A posthuman perspective on these phenomenal bodies emphasises humans as a plurality of ultra-social learners that come into being through different embodied experiences entangled in a material world. The processes moving us from novices to experts are not comparable to the learning processes of machines because we are humans in the plural, who consistently learn new values from each other. The difference between a novice and an expert only makes sense for humans, not for machines. Machines do not learn like humans to become intuitive cultural learners where meanings and materials constantly shift ground within phenomena.

Contrary to the Dreyfus brothers' original argument, in the posthuman perspective not even the novice can be said to learn rule-based like a machine. Our driver does not follow formal—value- or context-free rules to learn the skill of driving, even when she is a novice. The rules she learns to embody, and which are embedded in the instructor, in the car, in the road and in the signs along the road, are cultural—and learned through cultural learning processes. Whatever she learns as a novice is already entangled with what she has previously learned about cars, probably since she was a toddler. By taking a posthuman learning approach, we can conclude that when Dreyfus and Dreyfus referred to 'experience with concrete cases' as a driver of the process (Dreyfus and Dreyfus 1980, 5), this experience is about learning to become entangled in a community of drivers that collectively share a meaningful engagement with cars, gears and roads. Even if it involves constantly new intra-active entanglements, the process is also a process of learning as the novice becomes more and more culturally embedded in the community of drivers.

Here, potentially volatile humans and materials and their unruly, unpredictable behaviours are gradually handled in ways that lead her to be recognised as a more proficient member of the culture of driving. Thus, contrary to what the Dreyfus brothers claimed, there is no context-free learning going on at all. As this was originally the only possible link to the AI machine–learning community, even this possibility of machines learning like human novices is now severed. Humans do not learn context-free, but are culturally embedded in materialised and social collectives.

Will robots ever be able to learn like humans? If learning is considered posthuman, there are no 'context'-free activities. The intra-active workings of an instructor explaining some rules to a novice are already entangled with a material world that gradually has become meaningful through prior learning. What is meaningful is settled within phenomena. The phenomena of driving create novices and experts from within as the new driver begins to experience how actually driving as an 'advanced beginner' (Dreyfus and Dreyfus 1986, 23) takes shape. As she reaches competence, the intra-agency of her thinking body entangled in the phenomena of driving makes it possible to include other kinds of meaningful thinking, e.g. about how to find the fastest way to a destination, while her mindful collective body shifts gears in culturally suitable ways. In the intra-action, she sometimes experiences herself as unsure of the outcome of her actions, but the entanglement of materials always involves the unexpected. Upon reaching proficiency, the brothers introduced the interesting concept of 'holistic similarity recognition' (Dreyfus and Dreyfus 1986, 28). From the posthuman learning perspective, this entails that, within the phenomena of driving, a specific focus on 'driving' as a phenomenon is replaced by other kinds of embodied thinking and material engagements. As the driver learns to become 'intuitive', the prior learning of 'driving' is embodied as so much other prior cultural learning in her collective body within the phenomena of the world. What we can say from a Baradian point of view is that the phenomenon is no longer driving. Although the car is still part of the intra-action, the phenomenon shifts from 'driving' to, for instance, different sights along the road (e.g. cows in a field). The holistic similarity builds on former collective learning that makes 'cows' meaningful in different ways for different drivers (the driver might be a farmer or a desk clerk), but the driver's embodied placement in the car gives her a specific embodied experience of 'cows' as she drives by (e.g. no smells). For a farmer, her prior learning may still include prior learning regarding the smell of cows in the phenomenon. When she reaches stage five, she is not just fulfilling the process of embodiment, where 'driving' has become so intuitive that she no longer thinks about it, she is also now fully embedded in a cultural collectively of 'driving'. However, the materials are still unpredictable. What must be done is only 'simply done' if the cultural materials allow it.

Posthuman learning is a challenge for AI and machine learning because it changes our perception of 'the human' from a rational stand-alone individual to a collective of humans (in the plural) and non-humans. Posthumanism here refers to shifting perceptions of what constitutes humankind. However, in the disquieting theories of Barad, there is more at stake. Here we humans become posthumans because we are literally transformed into entanglements. Barad's line of thinking can in some ways easily be connected to the transformations proposed by some in the AI community (e.g. Kurzweil 2005), but she forgets that the mindful human body is always implicated in entanglements and that cultural learning will always help determine which boundaries can be established. We, therefore, need to restore the human-centeredness of entanglements (e.g. Gill 1991) in posthuman learning.

The posthuman learning I proposed is indebted to phenomenology, such as the embodiment proposed by the Dreyfus brothers, as well as to postphenomenology as proposed by Ihde, among others, as these perspectives make us aware of our shifting human–material relations with the world. The importance of the human perspective in our entanglements cannot be underestimated, but we need a new posthuman perspective on this 'human'. I do follow Barad in her dismantling of human exceptionalism—the view that humans are entitled to perform their rational cognitive transformations of the Earth. Barad does not specifically deny that human psychological processes can become entangled in phenomena; however, on the other hand she does not emphasise that prior human psychological processes (their learned languages, thinking and memories) are needed to meet the Universe halfway and perform phenomena. Phenomena can, as learned phenomena, not come into being without the prior cultural learning of humans.

From a Vygotsky-inspired learning perspective, I emphasise the collective mindful body involved in all intra-actions of phenomena. The body changes as the novice becomes an expert in a way that is not about an a priori abstraction, nor about simple cause and effect, but about culturally informed meaningful causes and effects that arise within phenomena of driving. Just as with the abstraction, causes and effects that constitute assumptions within the AI community are those that have been learned to be meaningful within the AI community. Machines do not learn like humans because they do not learn how to constantly make a volatile world meaningful, instead running on fixed abstract, yet material algorithms that are not about thinking, but about symbolic representation. Even in new machine learning, fixed parameters are in place (prefiguring inputs and outputs). The collective bodies of AI researchers and/or drivers are about a

gradually expanded, as well as ever-changing, spatial, ontological and epistemological collective embeddedness in a material world. What is machine and what is human is co-created and co-exists within phenomena. Nevertheless, when humans are involved, so is their prior learning; something that machines cannot replicate. Even if the AI community revolts against any notions of 'context'-free engagements, it is questionable whether machines will ever be able to learn like non-individualistic 'posthumans'.

References

Barad K (2007) Meeting the universe halfway: Quantum physics and the entanglement of matter andmeaning. Duke University Press, Durham

Boden MA (2006) Mind as machine: a history of cognitive science, vol 1. Oxford University Press, Oxford

Braidotti R (2013) The Posthuman. Polity Press, Cambridge

Dreyfus HL (1965) Alchemy and artificial intelligence. The RAND Corporation. Paper P-3244

Dreyfus HL (1979) What computers can't do. MIT Press, New York

Dreyfus SE (2004) The five-stage model of adult skill acquisition. Bull Sci Technol Soc 24(3):177–181. https://doi.org/10.1177/0270467604264992

Dreyfus S, Dreyfus H (1980) A five stage model of the mental activities involved in directed skill acquisition. (Supported by the U.S. Air Force, Office of Scientific Research (AFSC) under contract F49620-C-0063 with the University of California) Berkeley, February 1980. (Unpublished study). Washington, DC: Storming Media. https://www.stormingmedia.us/15/1554/A155480.html. Accessed 10 Oct 2017

Dreyfus HL, Dreyfus SE (1986) Mind over machine: the power of human intuition and expertise in the era of the computer. Free Press, New York

Ferrando F (2013) Posthumanism, transhumanism, antihumanism, metahumanism, and new materialisms. Differ Relat Existenz 8(2):26–32

Gill KS (1991) Summary of human-centred systems research in Europe, part 1. Systemist. J UK Syst Soc 13(1):7–27

Hasse C (2008) Postphenomenology—learning cultural perception in science. Human studies. Springer, Hamburg, pp 43–61

Hayles K (1999) How we became posthuman; Virtual bodies in cybernetics, literature, and informatics. The University of Chicago Press, Chicago

Ihde D (2002) Bodies in technology. University of Minnesota Press, Minneapolis

Ihde D (2003) Postphenomenology again. Working Papers from the Centre for STS Studies, nr 3, 2003, http://sts.au.dk/fileadmin/sts/publications/working_papers/Ihde_-_Postphenomenology_Again.pdf. Aarhus

Ihde D (2011) Of which human are we post? In: Gregory R, Hansell, Grassie W (eds) Transhumanism and its critics, Chapter 8. Metanexus Institute, Philadelphia, pp 123–135

Kurzweil R (2005) The singularity is near: when humans transcend biology. Viking, New York

Lyyra A (2015) Towards interaction machines. iSChannel 9(2):6–13

McDowell J (2007) What myth? Inquiry 50(4):338–351. https://doi.org/10.1080/00201740701489211

Minsky M (1986) The Society of Mind. Simon and Schuster, New York

Merleau-Ponty M (1962/1945) Phenomenology of perception. (C. Smith, Trans.). Routledge, New York, London

Mercer C, Trothen T (2014) The religion of technology: transhumanism and the myth of progress. In: Mercer C, Trothen T (eds) Religion and transhumanism: the unknown future of human enhancement. Praeger, Westport

More M, Vita-More N (2013) The transhumanist reader: classical and contemporary essays. Wiley-Blackwell, New York

Nagel T (1974) What is it like to be a bat? Philos Rev 83(4):435–450

Nath R, Sahu V (2017) The problem of machine ethics in artificial intelligence. AI Soc. https://doi.org/10.1007/s00146-017-0768-6

Sheets-Johnstone M (2000) Kinetic tactile-kinesthetic bodies: ontogenetical foundations of apprenticeship learning. Hum Stud 23(1):343–370 **(Kluwer Academic Publishers, Amsterdam)**

Vygotsky LS (1978) Mind in society: the development of higher psychological processes. In: John-Steiner V, Scribner S, Souberman E (eds) Cole M. Harvard University Press, Cambridge

Vygotsky LS (1987) Thinking and speech. In: Rieber RW, Carton AS (eds) The collected works ofL.S. Vygotsky, vol. 1 Problems of general psychology, Plenum, New York, pp 39–289.