



# The brain, the artificial neural network and the snake: why we see what we see

Carloalberto Treccani<sup>1</sup>

Received: 29 July 2019 / Accepted: 18 August 2020 / Published online: 12 September 2020  
© Springer-Verlag London Ltd., part of Springer Nature 2020

## Abstract

For millions of years, biological creatures have dealt with the world without being able to see it; however, the change in the atmospheric condition during the Cambrian period and the subsequent increase of light, triggered the sudden evolution of vision and the consequent evolutionary benefits. Nevertheless, how from simple organisms to more complex animals have been able to generate meaning from the light who fell in their eyes and successfully engage the visual world remains unknown. As shown by many psychophysical experiments, biological visual systems cannot measure the physical properties of the world. The light projected onto the retina is, in fact, unable to specify the physical properties of the world in which humans and other visually ‘intelligent’ animals behave; however, visual behaviours are habitually successful. Through psychophysical evidence, examples of the functioning of Artificial Neural Networks (ANNs) and a reflection upon visual appreciation in the cultural and artistic context, this paper shows (a) how vision emerged by random *trial and error* during evolution and lifetime learning; (b) how the functioning of ANNs may provide evidence and insights on how machine and human vision works; and (c) how rethinking vision theory in terms of trial and error may offer a new approach to better understand vision—biological and artificial—and reveal new insights into *why we like what we like*.

**Keywords** Human vision · Machine vision · Brain · Artificial neural network · Visual appreciation

## 1 Introduction

It is possible to consider metaphors as a way to attribute human characteristics to an animal, object or any other subject. At the same time, it is important to consider that abstract (non-physical) phenomena are understood through the attribution of physical features to the phenomena. For instance, the sentence *inflation rose in July* is interpreted using two concrete physical phenomena: inflation (an increase in size) and rising (a change in position) (Pinker 2013).

Metaphors are used in science and philosophy to explain the unknown, as well as a tool to generate new knowledge and provide a better understanding of many phenomena (Zarkadakis 2015). Every century, perhaps even every decade, has its own metaphors, and “when the use of a specific metaphor ceases and a new metaphor takes its place, we have

a ‘paradigm shift’ in the way science explains the world” (Zarkadakis 2015).

One main source of metaphors is the human brain and body. In the Book of Genesis Adam, for instance, is created out of dust and then life is infused into it—interestingly, the word human comes from the Latin *humus*, which means ground or earth. Later on, in the third century BCE, the invention of hydraulic and pneumatic systems provided a new paradigm to understand the human body as a “dynamically moving fluid within a mechanical body” (Zarkadakis 2015). In the sixteenth century, again, René Descartes described the human body as a complex machine, comparing muscles and bones to cogs and pistons.

The arrival of computer technology, however, changed the paradigm once more. The brain started to be compared to a computer, as it processes the information much similar to how a computer does. The computer–brain analogy perhaps finds its roots in the fact that information is transmitted

✉ Carloalberto Treccani  
carloalberto.t@my.cityu.edu.hk

<sup>1</sup> School of Creative Media-City, University of Hong Kong, Hong Kong, Hong Kong SAR

through electrical signals. Neurons,<sup>1</sup> like electronic components, indeed transmit information through a voltage change that much resembles the binary logic—0 and 1—used by computers.

From the 1950s, however, with the advent of Artificial Neural Networks (ANNs), this metaphor has begun to be revised. Unlike a computer, ANNs operate empirically, based on trial and error and on what worked best in the past, with no rigid rules or specific steps to follow—this distinction, which may seem to be of no particular importance, is fundamental to the understanding of this paper.

Using the new brain–ANN analogy, this paper introduces a new understanding of visual perception. Section 2 provides some essential information on the structure and functioning of the nervous system and briefly discusses its advantages. Important terminology is defined to better understand the entire text. Section 3 provides a basic understanding of the functioning of the ‘visual brain’ and introduces the idea that vision should be understood in empirical terms arising from trial and error during evolution and life-long learning. Section 4 provides a basic understanding of the functioning of ANNs. Section 5 clarifies the position of culture in visual understanding. Additionally, it provides insights into comprehend *why we like what we like*. Lastly, Sect. 6 presents some implications and possible consequences of ‘accepting’ the brain–ANN paradigm.

## 2 The nervous system

*Trichoplax adhaerens* is a marine animal roughly one millimetre in diameter without a nervous system (Senatore et al. 2017). However, despite this lack, as shown by Senatore and colleagues, *Trichoplax* is capable of a variety of behaviours typically found in animals with nervous systems. *Trichoplax* can track space, communicate through cellular transmission and exhibit different feeding behaviours: it can arrest its ciliary movement, used for locomotion, when algae are detected, showing that it has a sensory system able to detect nutrients and to communicate over short distances via chemical secretion. Plants also lack a nervous system but, like bacteria and protists, they can use environmental information to generate behaviours that enable them to survive and reproduce.

What are the evolutionary advantages of having a brain, then? Answering this question risks falling into some kind of hierarchical division, with organisms with a nervous system at the top of the pyramid of life and others at the bottom. It is certainly true that having a nervous system provides

numerous advantages. Nevertheless, having a brain is not a fundamental requisite for an ‘intelligent’ life form: most past and extant organisms are without a nervous system, and this absence does not seem to have caused them any particular problems.

No one knows when the first nervous system appeared; however, the need to survive in a constantly changing environment seems to have benefited those organisms ‘gifted’ with a nervous system (Purves 2019). The key distinction between non-nervous system organisms and those with a nervous system seems thus a *quantitative* distinction in the range of possible behaviours. The appearance of nervous systems and later on of central nervous systems enabled biological creatures to respond to the external environment in more sophisticated and useful ways (Robson 2020).

The nervous system is composed of specialised cells known as neurons and glia.<sup>2</sup> It is described for convenience as the central nervous system (brain and spinal cord) and the peripheral nervous system—both areas are of course in continuity. The task of the nervous system is to carry sensory information (e.g., heat) from the periphery to the brain to promote a behaviour (e.g., remove your hand). Each neurons signals via a bioelectrical signal, known as an action potential, which travels along the nerve cell and communicates the information to another neuron—the synapse.<sup>3</sup> Each cell remains independent and separate. In the case of vision, when photons stimulate photoreceptors cells in the retina, the sensory input is transformed into a neural signal and sent to specific areas within the brain. The information is then *processed*, and a behavioural response occurs.

Is certainly true that creatures gifted with *excitable cells* able to perceive and convey information about the outside world have evolutionary advantages compared to creatures without such cells. For instance, in the case of vision, the ability to create a visual representation of the world—e.g., identify a predator. Nevertheless, as previously stated, the majority of living beings that have existed and that currently exist lack a nervous system. Thus, having cells that can ‘represent’ the world and coordinate such representation with a behaviour (e.g., movement) should be seen as increasing the possibilities to representing the outside environment (Churchland 1989) rather than a fundamental requirement. Although the behavioural catalogue of different creatures varies tremendously, living organisms with and without nervous systems have developed strategies to pair sensory

<sup>1</sup> A neuron is a specialised cell for the transmission of electro-chemical signals.

<sup>2</sup> Glia cells are non-neural cells in the nervous system. Their main role is to provide structural support to neurons. They are not directly involved in the transmission of signals.

<sup>3</sup> Synapses are connections between neurons and a target cell. The role of synapses is to allow communication by receiving or transmitting chemical signals.

inputs with useful behavioural answers (Purves 2019). In summary, the advantage of having a nervous system seems that of having a richer behavioural repertoire.

### 3 Vision

According to Andrew Parker's *light switch theory* (2004), the change in atmospheric conditions during the Cambrian period and the subsequent increase in light has triggered the sudden evolution of vision and the consequent evolutionary benefits. Despite the fascinating idea proposed by Parker and the consequent belief of the supremacy of vision over other senses, this theory does not explain how organisms have been able to usefully pair a visual stimulus<sup>4</sup> (e.g., the 'image' of a prey) with a useful behaviour (e.g., eat the prey). After all, at that time, vision was a new and, therefore, unknown source of information. How have simple organisms and more complex animals been able to develop useful behaviours in response to visual stimuli, or put differently, how *visually gifted creatures*<sup>5</sup> have been able to create meaning from the light that enters their eyes?

The plain individual is probably convinced to see objects with a specific shape, because those objects have those certain physical features or to see them at a certain distance, because they are actually at that distance (Kanizsa 1997). However, as further explained in this paper, despite the apparent simplicity, how vision works, is still largely unknown—which is why, from their first appearance more than 50 years ago, computer vision systems remain to date, under certain circumstances, inaccurate, unreliable and easily deceived. Indeed, the need to further understand and replicate vision has only recently arisen, with the dream and need to build '*visually intelligent machines*' (Treccani 2018).

The conventional idea is that light carries with it information about the world that somehow resembles the world as it is (e.g., I see a house because the light input received by the sensory apparatus carries with it some 'houseness' information, such as shape). This belief is probably based on old theories of vision such as *intromission theories*, which see vision as light rays emitted by objects, and *extromission theories*, which instead understand vision as the emission of light rays from the eyes towards external objects. This idea is, however, incorrect, because light does not carry with it any information other than energy. Furthermore, the human perceptual apparatus cannot measure the physical parameters of the world, because it lacks the necessary instruments

and thus cannot retrieve the real properties of the world (as described more in detail later in this section). Besides, as the world cannot be assessed, even the idea of a representation of the world close enough to reality must be incorrect.

According to Parker's theory, the advantages appearing from the evolution of vision gave rise to numerous new animal behaviours. Seeing in colour, for example, is undoubtedly an immediate asset. As pointed out by Purves, "a visual system that can identify object boundaries based on the spectral distribution of light energy will, therefore, be more successful in responding to images" (Purves 2019). Here, Purves refers to the possibility of perceiving boundaries given by colours. Achromatic animals, for instance, cannot distinguish boundaries between two objects with a spectral difference but the same luminosity (i.e., brightness). Colours, a fundamental source of information that animals use to identify predators and poisonous plants or for reproductive purposes, are, however, a brain construction.

Colours are defined by variations in the wavelength (or frequency) of light. Red, for instance, has a wavelength of between 700 and 635 nm, whereas blue is between 490 and 450 nm. However, light wavelengths themselves do not correspond to any colours. The variation in the frequency of the wavelength, commonly understood as colour variation, is a vibrational variation (moving photons) in the amount of energy in a light wave. In addition, light waves that reach the retina always entail a combination of illumination, reflectance, and transmittance (Fig. 1a), and there is no analytical method to unravelling how these factors provide visually appropriate answers (Purves and Lotto 2003).

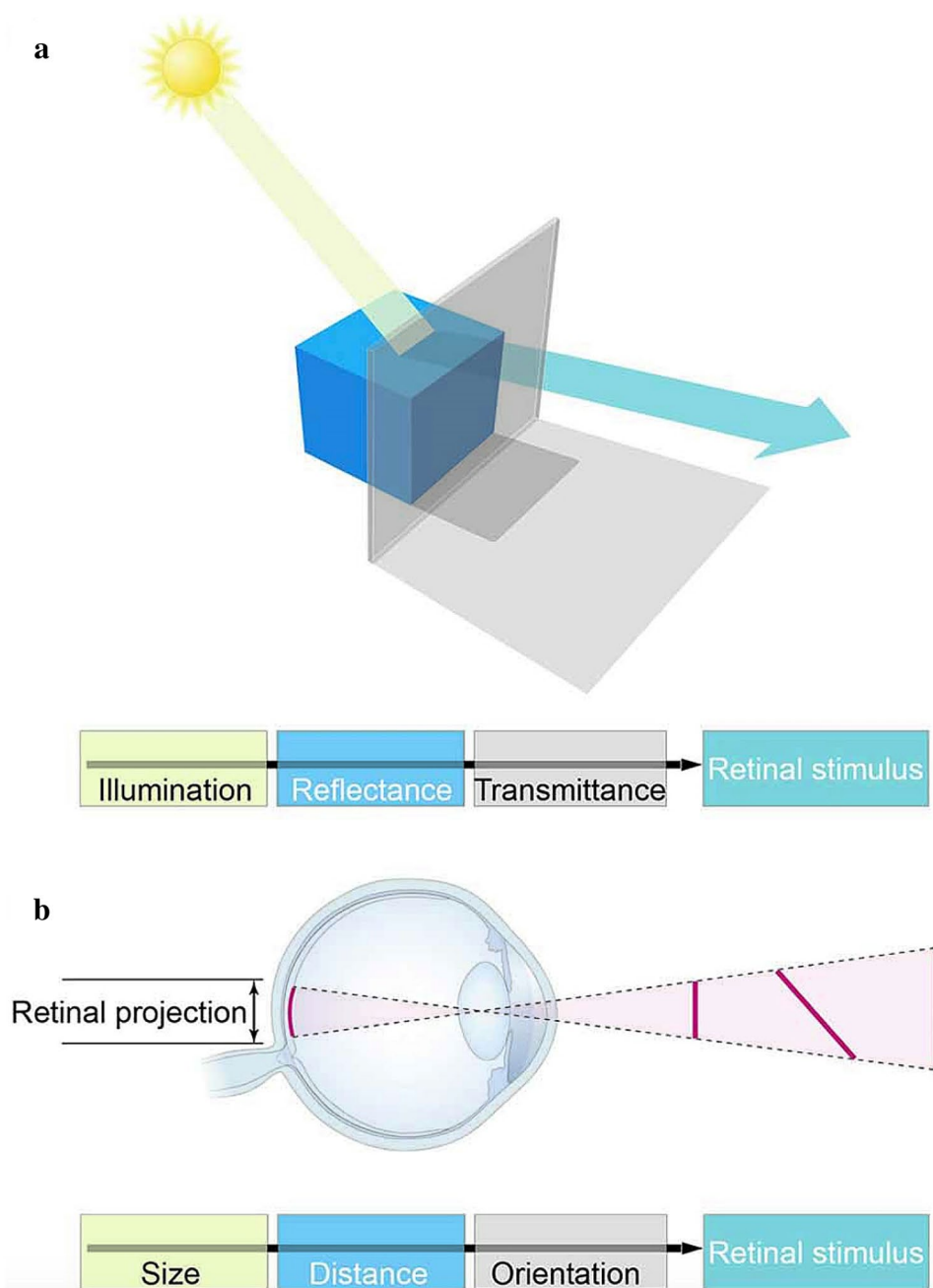
How do humans perceive colours? The capacity of the human eye to discriminate spectral variations is based on the sensitivity of retina cells to different light frequencies. Humans have two types of photoreceptor cells in the retina: rods and cones. Rods seem to play a minor part in colour detection. The three different types of cones are each characterised by a different type of photopigment. Each type of cone is sensitive to a different light frequency (i.e., colour). Yet this explanation does not seem to fully explain how and why colours are perceived. Additionally, the way geometrical properties are perceived shows the inaccuracy of the human vision in discerning reality once more. The human visual system, in fact, cannot retrieve the geometrical properties of the world. As shown in Fig. 1b, objects with different inclinations, sizes and at different distances can indeed generate the same retinal image. As human eyes, and more generally, the entire perceptual system, cannot retrieve the 'real' properties of the world, it is clear that visual perception must be a *generated perception* (Purves et al. 2014).

To explain the operation of visual perception, Dale Purves proposed the idea that vision should be understood in *empirical terms* "in which perceptions reflect biological utility based on past experience rather

<sup>4</sup> In biology an event, in the form of energy, that provokes and activates a receptor cell.

<sup>5</sup> Biological creatures *gifted* with photoreceptors able to transmit information about the outside world.

**Fig. 1** **a** The conflation of illumination, reflectance and transmittance. Many combinations of these objective parameters in the real world can generate the same values of luminance at the retina. **b** The conflation of physical geometry. The same image on the retina can be generated by objects of different sizes, at different distances from the observer, and in different orientations (Purves 2019).



than objective features of the environment” (Purves et al. 2015). Retinal images,<sup>6</sup> continue Purves “conflate the physical properties of objects, and therefore cannot be used to recover the objective properties of the world. Consequently, the basic visual qualities we perceive – e.g., colours, form, distance, depth and motion – cannot specify reality” (Purves et al. 2015). Visual perceptions, therefore,

<sup>6</sup> Images should be intended both as the light pattern detected by the retina, as well as the visual result of the processed light-pattern, meaning its visual representation—e.g., the image of a cat.

must emerge independently from any measurement of the world, as these measurements are not reliable. To paraphrase Purves, the perceptual information—the visual world—we experience, is determined by the frequency of light pattern<sup>7</sup> and its consequent importance in terms of survival. The association between the frequency of

<sup>7</sup> It is important to highlight that a light pattern does not resemble the physical world, nor any ‘image’ of the world produced by the human brain. A light pattern should be intended as the light configuration that is perceived, determined by its frequency of occurrence and its consequent usefulness (for more, see Purves et al. 2015., Yang and Purves 2004, Rao et al. 2002).

occurrence of a light stimulus (light pattern) and its consequent useful (successful) behaviour thus arises from trial and error during evolution and life-long learning.

If, as previously stated, animals' visual perceptual apparatus cannot assess the physical properties of the world, how are then images of the world formed? For Purves, if a given light stimulus occurs often, its value will be high. At first, different and random behaviours will appear in response to this stimulus. However, over time—evolutionary time and individual lifetime—an automatic link between the stimulus and the most successful behaviour will arise. According to Darwin and neo-Darwinian theory (the integration of Darwin's theory of evolution by natural selection and Mendel's theory of genetics as the basis for inheritance), when a mutation occurs and is randomly associated with a neural response that promotes survival/useful behaviour, this mutation and its consequent neural activity will tend to be passed to subsequent generations. In other words, if the image that arises from a given luminous-configuration and its consequent behaviour proves to be useful for evolutionary purposes, a link will be created. This successful creation will be disseminated to the next generation following evolutionary processes (thus, a non-useful behaviour will *not* tend to be passed to future generations).

Purves, even if referring here to the process of hearing, gives an explicative example that can be likewise transferred to vision: “This strategy works because it establishes an objective-subjective association in biological machinery that does not depend on the measurements of sound sources in physical reality. As before. The role of the physical world in this understanding of sensory neurobiology is simply an arena in which neural associations are empirically tested according to survival and reproductive success” (Purves 2019).

As humans perceive light “categorically—usefully but not at all accurately—because it is an extremely efficient way of perceiving visual stimuli that allow us to save brain cells to devote to the neuroprocessing of our other senses” (Lotto 2017), vision should be seen as a fast answer to a visual stimulus. Consider, for instance, the knee-jerk reflex. This reflex is simply an evolutionary ‘answer’ that establishes a successful behaviour. The activation of the nervous system, which is rapid and precise and without the need for any ‘brain computation’, promotes a response (i.e., extend the leg) that is useful (i.e., extend the leg to avoid falling if an object hits you). In this sense, vision should be considered a *reflex*: an automatic, quick and useful answer to a visual stimulus that does not require any measurement of the world (Purves et al. 2015). The role of vision, and more generally of the nervous system, is, therefore, to promote a biological advantage rather than to reveal *how the world looks like* (Purves 2019). In short, the role of vision is to promote what was *useful* to see in the past.

## 4 Artificial neural networks (ANNs)

Evidence of neuroscientific practices has been found in ancient societies all around the world. However, only with the advent of electronic technologies did scientists start to try to replicate the functioning of the brain. In the 1940s (following Turing's work in the 1930s), McCulloch and Pitts, respectively a neurophysiologist and a mathematician, began to investigate the possibility of neural computation. In 1943, they published a paper on the operational functioning of neurons and the construction of ANNs that could compute logical functions (McCulloch and Pitts 1943). With the idea of building a machine able to solve any logical operation, McCulloch first, with the help of Pitts later, embarked on the design of a mathematical model—a network—of brain functioning.

Towards the end of the 1940s, Donald Hebb published *The Organization of Behavior* (2002), in which he demonstrated that the more a neural pathway is used, the stronger it becomes. This concept is of fundamental importance: it shows that connections between neurons can change their *synaptic weight*, i.e., the strength of the connection. The stronger the connection, the more likely an association (stimulus-behaviour) will be to appear in the future. It is essential to understand this idea, as it is critical to the way humans and machines learn to see the world. In contrast, if seeing a stimulus—input—in a particular way shows itself not to be useful, it is less likely that the same neural pathway will appear.

In the past 10 years, ANNs seem to have gained the upper hand over algorithmic processing. Indeed, the “abilities to recognize patterns, make inferences, form categories, arrive at a consensus ‘best guess’ among numerous choices and competing influences, forget when necessary and appropriate, learn from mistakes, learn by training, modify decisions and derive meaning according to context, match patterns and make connections from imperfect data” (Greenwood and Bartusiak 1992) have made ANNs particularly successful—this is also why, all around the world, neuroscientists use ANNs to simulate the brain functioning. As example, a paper presented in 2018 by the DeepMind group (Google) (Silver et al. 2018) shows how the AlphaZero system beat the best human Go player in the world. Instead of using the *force* of an algorithm—following fixed procedures—AlphaZero learned, *tabula rasa*, how to win by playing against itself millions of time via a process of trial and error. By random trial and error (i.e., reinforcement learning), the system learns from wins and losses (i.e., empirical evidence) to adjust the parameters (i.e., synaptic weight) of the network, making it more likely to choose useful moves in the future. By ranking the frequency of winning moves, AlphaZero quickly learns how to become the strongest player in Go history.

In other words, differently from traditional game engines like IBM's Deep Blue<sup>8</sup> that rely on thousands of rules, fixed procedures and heuristic moves chosen by human experts, AlphaZero learns how to become the best player, *without in-built knowledge*—except the rules of the game—but by random self-play—trial and error (Hassabis et al. 2018).

Even if first created more than 50 years ago, it is only recently that ANNs have achieved satisfying results. ANNs are now used for image recognition tasks, to automatically identify objects, people, actions and places in a given image. Image recognition technologies are used by companies and governments to execute tasks such as guiding robots and vehicles, content searching in images, image labelling and medical diagnosis. Facebook, for instance, uses ANNs to help visually impaired users to identify people or objects in a photograph. Self-driving cars use ANNs to locate pedestrians and vehicles, and airports use facial recognition technologies as a biometric confirmation to allow entry to a country.

For the purposes of this paper, it is of great importance to clarify the driving principles of an ANN's functioning, as their understanding seems to provide insights into comprehend how the brain solve visual issues.

In his book *Intelligence Emerging*, Keith L. Downing explains the operating of an ANN as follows: “Take a brief pause from reality and imagine that you are the first-year coach of a professional basketball team. It is the first game of the season, and the score is tied with only seconds remaining. You call a timeout, consider your multitude of strategic options, and then decide to set up a play for the guy that everyone calls C. The play begins, C gets the ball, and though double-covered by two hard-nosed defenders, gets off a long shot. Swish! The ball goes through the net, you win the game, and C is carried off the court on the shoulders of his/her teammates. You have now learned some valuable information that will help throughout the season: C is a clutch performer” (Downing 2015). Let us also imagine a few other scenarios.

In a second case, a player known as D gets the ball and shoots, but the ball does not reach the basket, and the match is lost. The newspapers will attribute the defeat to D.

In a third scenario, as described by Downing, C passes the ball to B, who shoots and wins the match. In this case, both C and B will be glorified by the media and acknowledgement will be given to both—indeed, as Downing notes, basketball statistics include assists as evidence of a player's value.

In a fourth new scenario, “C passes to B who passes to (the guy everyone just calls) A, who makes the winning shot. A gets the points and the shoulder ride to the locker room, B gets the assist, but does C get (or deserve) anything? It could be the case that, before passing to B, C faked a pass or dribble attack that froze A's defender in place. This made it easier for A to come free to get the ball and shoot the winning basket. Such a contribution by C would not show up on a statistics sheet, though you may notice it. You may even praise C more than B or A afterwards in the locker room, since his/her fake-then-pass was obviously the key to the whole play. He set up a situation that then became routine for B and A” (Downing 2015).

After many games, it becomes clear that C plays a critical role during the final minutes of each game. In fact, the value of C is proved by the statistical significance of his performances. The proved value (i.e., usefulness) of a player during the closing minute of a match is essential to plan the strategy and organisation of the team through the rest of the season. Moreover, it is necessary to highlight that all shots, faults, mistakes, defensive impact, points, passes and all the events that precede the win are equally important when assigning value to a player or a particular team configuration. Was L fundamental for the winning of the game during his/her 2 min on the court? Was C a valuable player during the first part of the tournament but less valuable during the last part of the season? Was D a valuable player throughout the season despite the shooting error in the last game?

In short, the sum of all the *trials* and *errors* of the coach's choices and the consequent accumulation of experiences describe the essence of the functioning of an ANN—*reinforcement learning*.<sup>9</sup>

The more the trial and error search space is extended, the more likely is that a useful strategy will be found, although the incredibly large size of the search space and the difficulty of taking into account all of the possible variables do not allow the success of a decision to be determined with certainty. In the case of an *object recognition task*, for instance, it is difficult for a machine vision system to identify all the possible variables—e.g., size, shape, colour, orientation—of an object—e.g., a chair. Furthermore, an object can be partially hidden by other objects (a chair partially hidden by a table) or reflected into a mirror, creating even further difficulties (Treccani 2018). However, the ability of an ANN

<sup>8</sup> On 11 of May 1997, after a 6-match game, Deep Blue beat the world-best chess player—Garry Kasparov. The Deep Blue's win had an important symbolic significance in the advancement of 'intelligent' machine-artificial intelligence.

<sup>9</sup> Reinforcement learning, in machine learning's context, refers to the ability of an agent to learn by trial and error without supervision. Reinforcement learning, unlike supervised learning, does not require any labelled data. In the context of machine vision, a labelled data, for instance, might be a photograph in which the objects represented are specified—e.g., a cat, an apple or a house. Labels are usually obtained by asking annotators—humans—to make judgments about the content of a given image.

to successfully solve a task—for instance, detect every chair present in a given picture—increases over time. Increasing the exploration of the search space for the possible solution increases the possibility of a successful result.

Like Downing’s first-year coach at the season’s start, an ANN begins with a series of random decisions (e.g., to choose player D instead of C)—exploration. Over time and through numerous attempts, the network uses the information gathered to craft more useful decisions—exploitation. As the ANN explores the space of possible choices, it learns that certain decisions or actions lead to a reward while others lead to negative consequences.

In short, an ANN solve *the problem of vision* by millions of random trial and error or, in other words, through a well-indexed database of past experiences and not through logical procedures. Just as an ANN system learns how to recognise faces among millions of other faces, distinguish different dog breeds or recognise and describe a scene, a biological system learns how to see by countless trials and errors during evolution and individual lifetimes. In light of this, a new, *wholly empirical understanding* of the way humans and machines see based on trial and error is therefore needed (Purves et al. 2015).

## 5 The snake

The idea that vision emerges empirically, by trial and error, during evolution, life-time experience and training period—in the case of an ANNs—was previously presented in Part II and III. Extending this hypothesis to the cultural and aesthetic realm, part IV, will suggest that also visual appreciation should be understood as a useful behaviour emerging by trial and error.

Culture plays an important role in visual understanding. Nevertheless, culture has to be understood in evolutionary and biological terms, which means emphasising the evolutionary advantages of developing cultural traits—ideas, technologies and behaviours—that can be transmitted from an individual to another via parenteral and social learning. This idea implies that culture can be seen as a biological extension or, as noted by Creanza, as “the extension of biology through culture” (Creanza et al. 2017).

Relevant in this regard is the case of *snake detection theory*, which holds that “humans and other primates can detect snakes faster than innocuous objects” (Van Le et al. 2013). In *The Fruit, The Tree, and The Serpent: Why We See So Well*, Lynne Isbell argues that “When snakes (the Serpent) appeared, a particularly powerful selective pressure... favored expansion of the visual sense” (Isbell 2009). Isbell argues that the environmental pressure caused by the appearance of snakes—as competitors and predators—and their threat to survival was a possible trigger for the complexity of

the primate visual brain, its enlargement and the particular sensitivity towards snakes. “Across primate species, ages, and (human) cultures, snakes are indeed detected visually more quickly than innocuous stimuli, even in cluttered scenes. Physiological responses reveal that humans are also able to detect snakes visually even before becoming consciously aware of them” (Van Le et al. 2013).

Interesting is the reference that the author makes to the events of the Garden of Eden described in the Old Testament. Eve’s mistake was, in fact, noticing the snake. If she had not noticed the animal, she probably would not have eaten the apple. Snake references are found not only in Judeo-Christian confessions but also in other religions and cultures. Snake representations are found in pre-Christian societies, on Sumerian amulets, Iranian boxes, Greek and Chinese mythology. The presence of snakes in different cultures and the fear of snakes (ophidiophobia), seems thus to have an evolutionary explanation—a visual reminder that snakes are dangerous. As Isbell noted, “ophidiophobia may go way, way back, to at least 30–35 million years ago when the first Old World monkeys and apes, the so-called catarrhine primates, are thought to have appeared. Ophidiophobia may even extend farther back to 60 million years ago when the first generalised simian primates, the anthropoids, are thought have appeared. If so, this timeline might help explain the shared ophidiophobia of all anthropoids, including humans” (Isbell 2009).

The enlargement of the visual brain in primates, for Isbell, seems to have provided a fast and *automatic* ‘predator detection system’ (Isbell 2009). “The ancestral environment of primates uniquely affected them to link vision with automatic, fast, accurate, and adjustable reaching and grasping, and to improve upon vision as a way to detect and avoid predators” (Isbell 2009). As also pointed out by Purves (see Sect. 3), Isbell seems to refer to vision as an *automatism*. The advantages of understanding vision as a *reflex* are in fact: to provide a fast, direct and accurate response to the external environment that does not require any computation or processing, as the ‘computation’ have “already been accomplished by laying down connectivity instantiated by feedback from empirical success over evolutionary and individual time” (Purves et al. 2015). The capacity of humans to visual discriminate an object, for instance, is in fact in the order of tens of milliseconds—this may also provide insights on why, when looking a particular painting, for instance, we are immediately captured by it.

Snakes visual sensitivity and ophidiophobia thus appear to be an evolutionary aid: those who did not respond to snakes with a proper behaviour (e.g., running away) would have fewer chances to survive compared to animals with a more useful behavioural answer (Isbell 2009). The presence of snakes in human artefacts and religions can be explained through the behavioural advantages that arise from the

ability to visually detect snakes. In other words, a particular *visual sensitivity* towards snakes seems to have had cultural repercussions that justify the presence of this animal and its *visual appreciation*<sup>10</sup> in artistic productions all around the world.

It is certainly important to understand the value of culture in the transmission of useful behaviour, for instance, of techniques related to the construction of a tool. It is well known that several non-human species exhibit cultural transmissions. Chimpanzees and macaques, for instance, can build tools like hammers to open nuts using stones. However, even the value of transmitting visual appreciation skills should not be underestimated, it should be read instead as an evolutionary aid in the same fashion of a chimpanzee's ability to build a hammer, although with more degrees of separation.

As shown by Michael Baxandall in *Painting and social experience in fifteenth century Italy* (1988), having the ability to appreciate a painting was a necessary social skill to master for the upper middle and aristocratic Italian classes of the Renaissance. Possessing these *visual skills* gave access to a series of benefits in the form of social connections. In this fashion, artistic visual competence can be seen as an evolutionary aid. These *visual capacities*, although not immediate, must be understood as useful answers to the environmental and social 'selective pressure'. Every given period, in fact, has its own ecological 'selective pressure' and its consequent *useful visual behaviour*.

To trace the biological value of visual appreciation is surely complicated. However, exploring in this direction can reveal meaningful insights into *why we like what we like*. Furthermore, understanding vision and visual appreciation arising from trial and error during evolution and life-long learning can provide a new understanding to study human perception and culture.

## 6 Conclusion

The idea that the visual brain may operate like an ANN, and that vision works in empirical terms poses considerable difficulties both in the science and the humanities; however, as previously shown, there is much evidence that justifies this new analogy. Reconsidering vision in terms of trial and error implies the necessity to revise some of the ideas proposed in this regard, particularly in the fields of psychology and theory of perception studies. However, visual perception issues also need to be rethought in the perspective of brain–ANNs

analogy, since it seems to provide some insight into how human see and visually understand the world.

As demonstrated in Sects. 3 and 4, there are clues that suggest that vision emerges empirically as an automatic answer to a stimulus. The link of the frequency of a light stimulus and consequent useful behaviour determines what is seen. In the same fashion, by exploring all the possible behaviours (e.g., actions, choices) and changing the relative synaptic weight until the best 'move' is found, ANNs successfully learn how to solve a visual task (e.g., object classification, scene classification and image segmentation), possibly as biological creatures have done during evolution. The analogy should then be clear: the way humans and other animals build their way to visually understand the world closely resembles the operational functioning of an ANN—and vice versa.

This idea may seem suspect to many and generate a certain antipathy in others, but the possibility of exploring the vision (and, more generally, the functioning of the whole brain) in terms of trial and error can provide a first model of *why we see what we see*, to date still missing. Although the theories of the last century, especially in the field of psychology (such as Gibson's bottom–up theory and Gregory's top–down theory) and later in neuroscience (such as Marr's computational theory), had the merit of providing further elements of understanding of how the visual system may work. Likewise, these theories were not able to provide a clear explanation of *how*, and most importantly, *why* the 'visual world' is constructed (i.e., the relation between the physical properties of the world, the visual stimulus that falls onto the eyes and the consequent mental image).

As previously discussed in this paper, the physical properties of the world cannot be measured, and the degree of discrepancy between reality and perceived reality must be substantial. Thus, the idea that a subject is able to respond to a stimulus, based exclusively on the features of the stimulus itself (bottom–up theory) appears insufficient. Furthermore, the idea of seeing as knowing (top–down theory) seems equally insufficient, since the state of the world is unknown. However, by random trial and error, by ranking the frequency of the appearance of a light stimulus, the success of the response to that particular stimulus and the consequent neural wiring, animals, and more recently machines, seem to have been able to create a useful visual representation of the world. The role of vision then "is not to reveal the physical world, but to promote useful behaviours" (Purves et al. 2015).

In trying to understand visual perception and its guiding principles, it has to be clear that the human brain evolved from earlier brains, and that its capacity and limitations have a historical basis (Churchland 1989). The selective pressure that developed visual perception—and more generally, all perceptions—did not arise rationally and logically but

<sup>10</sup> Visual appreciation refers not only to the aesthetic appreciation of a work of art but also to the ability to analyse, describe, interpret and make connections between works of art and their cultural context.



instead from the need of living beings to successfully deal with the surrounding world. As the world cannot be conquered by the visual sensory system, animals—including humans—have learned to *usefully represent* and respond to the physical world through trial and error during evolution and individual learning. As an ANN learns how to solve a visual task by randomly trying the possible ‘moves’, so the brain has learned how to successfully respond to the world outside itself.

Comprehending how visual understanding was won by simple organisms first, more complex animals later and—partially—by machines recently, may be of great help as it may help to unhinge wrong assumptions and test the validity of new ones. Abandoning the nest of common-sense conceptions about vision—both biological and artificial—will certainly lead to further complications that nevertheless deserve to be investigated to better understand and provide a first, albeit imperfect, answer to the relationship between *what we see and what we know* (Berger 2008).

**Acknowledgements** Many thanks to Olli Tapio Leino for his continuous support, to Patricia Armati for her comments and to Dale Purves for his guidance in writing this paper.

**Funding** The writing of this paper was possible thanks to the financial support of the School of Creative Media-City University of Hong Kong and the University Grants Committee, Hong Kong, SAR, China.

## References

- Baxandall M (1988) *Painting and social experience in 15th century Italy*. Oxford University Press, Oxford
- Berger J (2008) *Ways of seeing*. Penguin Classics, London
- Creanza N, Kolodny O, Feldman MW (2017) Cultural evolutionary theory: how culture evolves and why it matters. *Proc Natl Acad Sci* 114(30):7782–7789. <https://doi.org/10.1073/pnas.1620732114>
- Churchland PS (1989) *Neurophilosophy toward a unified science of the mind-brain*. A Bradford Book, Cambridge (MA)
- Downing KL (2015) *Intelligence emerging. Adaptivity and searching in evolving neural systems*. Intelligence emerging. MIT Press, Cambridge
- Greenwood A, Bartusiak MF (1992) *Neural networks: computational neuroscience: a window to understanding science at the frontier takes you on a journey how the brain works?*. The National Academies Press, Washington, Science at the Frontier
- Hassabis D, Hubert T, Schrittwieser J, Silver D, et al (2018) AlphaZero: shedding new light on chess, shogi, and Go. DeepMind Blog. [https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-andgo?fbclid=IwAR0j8BhGOWaLBPRu00pVBEX0g4TgztWwTxA4\\_J3ozYsBeymnG5\\_QQWoqMXg](https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-andgo?fbclid=IwAR0j8BhGOWaLBPRu00pVBEX0g4TgztWwTxA4_J3ozYsBeymnG5_QQWoqMXg). Accessed 28 January 2020
- Hebb DO (2002) *The organization of behavior: a neuropsychological theory*. Psychology Press, London
- Isbell LA (2009) *The Fruit, the tree, and the serpent: why we see so well*. Harvard University Press, Cambridge
- Kanizsa G (1997) *Grammatica del vedere Saggi su percezione e Gestalt*. Il Mulino, Bologna
- Lotto B (2017) *Deviate: the science of seeing differently*. Hachette Books, New York
- McCulloch WS, Pitts WH (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5:115–133
- Parker A (2004) *In the blink of an eye*. Basic Books, New York
- Pinker S (2013) *Stylish Academic writing*. <https://www.youtube.com/watch?v=IE-TTz13P7w>. Accessed 18 January 2020
- Purves D, Monson BB, Sundararajan J, Wojtach WT (2014) How biological vision succeeds in the physical world. *Proc Natl Acad Sci USA* 111(13):4750–4755
- Purves D, Morgenstern Y, Wojtach WT (2015) Perception and reality: why a wholly empirical paradigm is needed to understand vision. *Front Syst Neurosci* 9(November):1–10. <https://doi.org/10.3389/fnsys.2015.00156>
- Purves D, Lotto B (2003) *Why we see what we do: an empirical theory of vision*. Sinauer Associates, Sunderland MA
- Purves D (2019) *Brains as engines of association: an operating principle for nervous systems*. Oxford University Press, Oxford
- Rao RPN, Olshausen BA, Lewicki MS (2002) *Probabilistic models of the brain: perception and neural function*. MIT Press, Cambridge MA
- Robson D (2020) A brief history of the brain. NewScientist. <https://www.newscientist.com/article/mg21128311-800-a-brief-history-of-the-brain/>. Accessed 26 December 2019
- Senatore A, Reese TS, Smith CL (2017) Neuropeptidergic integration of behavior in trichoplax adhaerens, an animal without synapses. *J Exp Biol* 220(18):3381–3390. <https://doi.org/10.1242/jeb.162396>
- Silver D, Hubert T et al (2018) A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 362(6419):1140–1144. <https://doi.org/10.1126/science.aar6404>
- Treccani C (2018) How machines see the world: understanding image annotation. *NECSUS. Eur J Media Stud* 7(1): 235–254. <https://doi.org/10.25969/mediarep/3425>
- Van Le Q, Isbell, et al (2013) Pulvinar neurons reveal neurobiological evidence of past selection for rapid detection of snakes. *Proc Natl Acad Sci* 110(47):19000–19005. <https://doi.org/10.1073/pnas.1312648110>
- Yang J, Purves D (2004) The statistical structure of natural light patterns determines perceived light intensity. *Proc Natl Acad Sci* 101(23):8745–8750
- Zarkadakis G (2015) *In Our Own Image*. Rider Books, London

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