Studi di estetica, anno LIII, IV serie, 1/2025
ISSN 0585-4733, ISSN digitale 1825-8646, DOI 10.7413/1825864700

Guido Bitossi

Further steps towards an enactivist pragmatist theory of picture perception

Abstract
How does picture perception work? This paper suggests the outline of an enactivist and pragmatist account of picture perception. After discussing representationalist interpretations of the phenomenon, I claim that picture perception is a form of visuomotor reenactment scaffolded by the picture. In the last part of the paper, I show how this account practically works. This hypothesis might be a theoretical framework for future empirical research.

Keywords
Picture perception, *Enactivism*, *Reenactment*

Received: 31/01/2024
Approved: 06/09/2024
Editing by: Giulia Milli

guido.bitossi@uniroma3.it (Università degli studi di Roma Tre)

Introduction

Gabriele Ferretti (2018; 2021) has developed a neural account of picture perception that updates the *dorsal/ventral account of picture perception* (from now on, *DVAPP*) of Bence Nanay (2010; 2011; 2015). The premises of both accounts are in the theory of *seeing-in* (Wollheim 1987). With the term *seeing-in*, we refer in literature to that special perceptual experience, which is the experience of figurative pictorial art. This experience has a two-fold quality: it is composed of a *configurational fold* (the perception of the surface of the picture) and a *recognitional fold* (the perception of the object depicted). Nanay's *DVAPP* suggests that these two folds are related to two different brain areas: the *dorsal stream* of vision (also known as *where stream*) and the *ventral stream* of vision (also known as *what stream*)[1]. For Milner and Goodale (1995), these two visual areas process two different features of visual perception: the first one controls motor execution, while the second one activates in cases of object recognition. Ferretti wants to update the *DVAPP*, drawing on new empirical results that suggest that the dorsal and the ventral stream are not two separate pathways but rather two networks in constant interplay (Verhoef *et al.* 2011; Perry, Fallah 2014).

I think that with the fast development of enactivism, it is time to suggest a more embodied account of how picture perception works. Indeed, the *DVAPP*[2] is inconsistent with the tenets of what is broadly called embodied cognition, not to say of enactivism. First, because the *DVAPP* is a theory of a strong internalist flavor. In one of his latest papers, Ferretti (2021) declares that action is essential to picture perception because it helps the beholder discriminate a picture from trompe l'oeil and ordinary perception. However, the role he gave to action seems

[1] According to Nanay, the ventral stream can process both the recognitional and the configurational fold in rare cases. It is in these cases for Nanay that *pictorial inflection* occurs. The *pictorial inflection* (for a review, see Hopkins 2010) occurs when the object depicted in the *recognitional fold* shares a class of properties (called *inflected properties*) with the *design* (a subset of the brush strokes composing the *configurational fold* that permits the recognition of the object depicted) of the picture.
[2] With the acronym DVAPP (without other specifications), I refer to both Ferretti's and Nanay's accounts since their theoretical core is the same. When dealing with single aspects of a specific account, I will explicitly refer to Ferretti's papers (in case I focus on his theory) or Nanay's papers (in all the other cases).

unsatisfying from an enactive view. Secondly, the DVAPP endorses a form of the sandwich model of the mind (Hurley 1998), which aligns with first-generation cognitive science. Indeed, the *DVAPP,* particularly in Ferretti's version, argues that picture perception is always guided by the top-down decisions of the mainly ventral interplay, compared to which the mainly dorsal interplay has almost a secondary role. This account has no place for bottom-up influences from the body to the brain.

Given these weaknesses, I want to suggest a different conception of picture perception inspired by enactivism and pragmatism. According to this conception, picture perception is primarily guided by perception and movement in a process of visuomotor reenactment. Neural representations leave the stage to a visuomotor process of interaction run from the beginning by bodily movement. I suggest that picture perception is a bottom-up process rather than top-down, for which picture perception starts from the bottom (a flexible and intelligent visuomotor know-how), reaching only subsequently the top (know-that) and only to attune and refine the *know-how*.

This account is an externalist account instead of an internalist because it acknowledges, drawing on the enactivist tradition (Varela *et al.* 1991; Noë, O'Regan 2001; Noë 2002; Gallagher 2005; Gallagher 2017; Hutto, Myin 2012), the constitutive and pre-noetic role that the body and the environment play in creating cognition. It is inconsistent with the sandwich model because it rejects a rigid dualism between cognitive control and motoric processes in line with the *meshed architecture* model (Christensten *et al.* 2016).

Finally, it is a pragmatist account because it is grounded in an anti-autonomist conception of the aesthetic experience (Dreon, Vara Sanchez 2022), drawing on the tradition of pragmatist aesthetics (Dewey 1989; Shusterman 1992). The core of the aesthetic experience of a picture is not inside the head but rather outside of it, in a flexible and intelligent bodily interaction with the work of art.

1. *For an externalist account of picture perception*

As I have said before, Ferretti's account has a place for action (Ferretti 2021). But first, for him, action comes into play only at the end of picture perception, and second, it only concerns the picture's surface. For Ferretti, the mainly ventral interplay cannot start a process of *action planning* and *response selection* when inspecting the content depicted. The

contrary happens when the mainly ventral interplay is inspecting the surface. It is only in this last case that a process of interaction starts. Ferretti explains to the reader that the surface (like all tridimensional objects) possesses a property known in the literature as the plastic effect (Vishwanath 2011) that automatically triggers action planning, while the depicted object, with its lack of depth, does not. To clarify, the bodily interaction caused by the action-planning process is the final product of a representational neural process. Action is for Ferretti only a mechanical *know-how* process that blindly follows the orders from the top of the intelligent *know-that* of the *mainly ventral interplay*. We can conclude from this general outlook that for the *DVAPP* the core of picture perception happens in the brain. Even if it is true that in Ferretti's account (Ferretti 2021), motor execution is what permits the beholder to differentiate between picture perception and *trompe l'oeil* perception, it could be objected that Ferretti here is not talking about action (as he claims), but instead of certain neural networks (the mainly dorsal interplay) to which real bodily action is dependent and secondary.

On the contrary, I want to suggest that the visuomotor interaction process happening outside the brain plays a role of primary importance in picture perception and that it is triggered not only by the surface properties but also by the content depicted. Picture perception is a *know-how* and a *skilled behavior* that consists in a visuomotor interaction. This process cannot occur immediately but extends over time. In this paper I argue for an externalist conception of picture perception that can be understood more easily if we look at the recent development of enactivism. For enactivism, the basic unit of cognition is not the central nervous system but a whole composed of the brain, body, and environment.

Gallagher (2017) clearly expresses the shift that has taken radical enactivism far from traditional cognitivist views. In particular, he states that cognition results from a holistic interaction between the brain, the body, and the environment. The *DVAPP*, instead, is still consistent with the *methodological solipsism* of first-generation cognitive science. This philosophical position, first adopted by Jerry Fodor (1980), argues that the functional role of mental states should be studied in relation to the neural states in which they are implemented. The concept of function, then, is restricted to the function the mental state has in the inner system of the mind-brain. The *methodological solipsism* is also present in the work of linguists such as Chomsky (1995). In the 80s, philosophers like Gilbert Harman (1988) advocated for a form of "wide functionalism" in which the

functional role of mental states is related to external factors. In this case, the function of a mental state refers to parts of the environment or of the body. For example to certain biological features of the environment that select certain functional behaviors because more winning. Enactivism endorses a similar externalist position against the internalism of Fodor and Chomsky. For enactivism: "the explanatory unit of cognition (perception, action etc.) is not just the brain, or even two brains in the case of social cognition, but dynamic relations between organism and environment, or between two or more organisms, which include brains, but also include their own structural features that enable specific perception-action loops, which in turn effect statistical regularities that shape the structure and function of the nervous system" (Gallagher 2017: 11).

There is evidence in neuroscience that the environment, the body, and the brain aren't three separable entities but rather *interaction-dominant* systems (Kelso *et al.* 2013; Van Orden *et al.* 2003; Dotov *et al.* 2010; Bedia *et al.* 2014; Shockley *et al.* 2003; Riley *et al.* 2011). The *interaction-dominant* system works differently from the *near-decomposability* systems of first-generation cognitive science. In an *interaction-dominant* system changes in the body or in the environment can directly alter the behavior of the brain and vice versa: the whole system is more than the simple sum of the parts because each change in one component determines unpredictable changes in the other two.

Regarding the body, Gallagher writes (Gallagher, Zahavi 2007) that our structural external features have a pre-noetic role in shaping cognition. This does not simply mean a causal influence from body to brain, as Adams and Aizawa (2008) stated, but rather a constitutional role of our flesh in creating cognition, according to a concept of cognition not synchronic, but rather diachronic (Kirchhoff 2015). This appears immediately true if we think at how bodily features shape our perceptual world, giving us the possibility of identifying affordances in the perceptual field (Gibson 1979) that depend structurally on the design of our legs, arms, eyes, and so on. This constraint is so strong that it is possible to argue that the worldview of a mosquito or of a dog consists of an *Umwelt* (Von Uexküll 1934) completely different from ours, given the different body that is proper of each of these creatures.

If this is true, picture perception must also be a holism. In this holism, the body reacts actively to what is seen in a process of reenactment, with the aim of recognizing and giving an imaginary reality to what is on the canvas. The brain, which is included in the complex process of bodily adjustment with the picture, is also attuned to all the subtle changes that

happen in the body because the body's structure shapes inextricably the brain and vice versa. Sometimes, in this attunement, the brain can positively contribute to the adjustment. As we will see, when interaction encounters problematic cases, the brain could contextually point with cultural knowledge to a possible solution, re-adjusting the relation between body and painting.

What are the features of this bodily interaction? A good part of this interaction is put into action with our eyes. Research around eye movements has become possible with the invention of eye-tracking technologies. After this invention, new important studies on the inner workings of our visual system were conducted (Yarbus 1967). Many recent studies have investigated the relationship between eye movements and language production. Some also use images (Johansson *et al.* 2006).

Johansson, Holsanova and Holmqvist, in particular, have shown in four experiments that: eye movements tend to follow the spatial position of objects told in a story; eye movements tend to follow the same spatial relations when the story is recalled verbally; eye movements tend to follow the spatial relations seen in a picture during a spoken description; and they prove that these results happen both when the experimental subject is in front of a blank whiteboard or in complete darkness. Finally, they list several hypotheses as possible explanations for these results. Two are the most interesting. The first is the *pictorial theory* of visual imagery of Kosslyn (1994). This is the hypothesis that Johansson, Holsanova, and Holmqvist find more convincing. Kosslyn argues that eye movements are a clue that testifies to the presence in the head of a process of *internal image representation* that is simultaneous to every linguistic performance. This hypothesis is based on other empirical research (Kosslyn *et al.* 1995) that shows how visual imaging and visual perception are based on the same neural networks. However, there are also more enactivist hypotheses that can explain the phenomena.

The principal one is based on the concept of *reenactment* (Thomas 1999, 2010; Thompson 2007). For the proponents of reenactment, eye movements in the absence of stimuli are the result of a process of visual imagery, as for Kosslyn. But differently from what is stated by the pictorial theory, enactivism states that imagery does not consist in visual representations in the head, but in the repetition of ocular patterns highly similar to those elicited by the presence of the situation imagined. This theory, proposed first by Thomas (1999), is an extension of the *perceptual activity theory* (Hochberg 1968) and of the *active vision* approach (Bajcsy 1988). These two accounts see perception as a continuous refinement of

schemas with the aim of probing and exploring the environment. As Foglia and O'Regan (2015) pointed out, enactive imagery should not be confused with the sensorimotor approach to imagery, for which imaging consists simply of recalling the sensorimotor know-how relative to the object imagined. Enactive imagery for being enactive should always be concretely executed through a particular behavior, that is, the same behavior the subject would have enacted when in front of the real stimulus. According to the reenactment thesis, eye movements follow the spatial terms in language description because the speaker re-enacts the same saccades and fixations he would have enacted in front of the visual scene described, and through this reenactment, he actively imagines the situation that he is verbally referring to. Thomas (1999) argues that reenactment theory can be a helpful tool for explaining not only the concept of imagery but the broader concept of imagination and seeing-as. A child *sees* a doll *as* a real baby because she reenacts the same perceptual behavior she would have enacted with a real baby. However, the perceptual feedback the child gets from a doll differs significantly from the perceptual feedback she could get from a real baby. Said otherwise, the concept of reenactment can be applied in its original formulation not only to pure imagination but also to situations in which the distinction between perception and imagination is more subtle. It is easy to see that the observation of pictures is one of these situations.

I want to argue that the interaction process in front of the picture I mentioned consists of a visuomotor reenactment. I hypothesize that two kinds of pictorial reenactment can be possible, although my hypothesis should be the object of experimental investigation. In the first case, the beholder reenacts in front of the content depicted x specific eye movements to recognize what is depicted. These movements are only used during picture perception and have been previously learned from observing pictures. In the second case, the beholder reenacts the same eye movements he would have enacted in front of the real x in front of the content depicted x. These movements help the beholder give imaginary reality to what is depicted and *see* the blots *as* something else. When we look at the duck-rabbit illusion, we see the illusion as a duck simply because we move our attention and gaze to the left side of the picture, towards the beak. We put in motion the same eye movement we would have enacted when recognizing the real head of a duck. The same could be said when we *see* the illusion *as* a rabbit. In this case, our attention is directed towards the right side of the picture, towards the mouth of the rabbit. The reenactments are composed of two elements: on

the one hand, the motoric aspect, and on the other hand, the perceptual (or sensory) one. The two are closely linked in what is a complex coordinated action.

It should be noted that the reenactment could not be possible without the presence of the picture, which plays the role of a material scaffold for this process. Pictures are not objects that do not play a part in the recognition process; instead, they constitute as much as our eyes and our bodies the cognitive process of picture perception itself.

Does looking at a picture mean building mental representations in our heads? Or rather reenacting a series of visual schemas in front of it? Enactivism has traditionally been critical of the use of first-generation cognitive science of representations. Enactivism argues that explaining complex processes using representations means necessarily falling into a *homuncular regress*. Indeed, a representation always needs an observer or interpreter, who, on its part, needs to interpret with the help of another representation and so on. For enactivism, we should never use representations (as explanans) to explain other representations (as explananda). Instead, the right theoretical move is to define higher representational abilities using more simple coupling processes between action and perception.

It could be argued that there is no reason to think that if ordinary perception implies action, then necessarily picture perception implies action. The remark contained in the paragraph above shows us that this line of reason is understandable but ultimately false: given that pictures are representations and that enactivism rejects explanations in which representations are explained through the use of other representations (Di Paolo *et al.* 2017), it is necessary to conclude that picture perception should be explained through more simple behaviors (movement, perception). The alternative otherwise is the *DVAPP*, but the risk of *homuncular regress* in this case is highly probable.

To resume my account, I argue that the body and the picture are engaged in a process of interaction that consists of a visuomotor reenactment. This reenactment causes changes in the brain, which, on its part, is attuned to the adjustments between the body and the picture. Sometimes, when the beholder is in front of problematic cases, abstract cognition can help restart the process of reenactment that previously stopped.

2. *Against the sandwich model*

However, the interaction process is not only visual. It is attuned and refined more and more precisely thanks to the action of two other elements closely intertwined with perception: the first one is the sensorimotor system, and the second one is abstract cognition. The mind's inner workings for classical cognitivism are characterized by their inflexibility. At first, the brain perceives, input data are sent to a central processor, and finally, the central processor moves the body and begins the action. Susan Hurley (1998) referred to this schema when she coined the famous *sandwich model metaphor*. John Dewey, in his equally notorious article *The reflex arc concept* (1896), criticized the psychological theories of his time with an analogous argument, questioning their interpretation of the mind with his criticism of what he called the *reflex arc concept*. In Dewey's model, perception, action, and cognition are not cognitive processes ontologically fixed from the start, but elements characterized by a functional value defined by their interrelation. Dewey's theory is very similar to today's embodied models of cognition.

The *DVAPP*, at least in Ferretti's formulation, is coherent with Hurley's sandwich model. With the help of perceptual inputs, the brain creates representations thanks to the mainly dorsal interplay and the mainly ventral interplay. Then, the action planning and response selection process starts when the mainly ventral interplay recognizes the plastic effect in the object of perception (as it happens for surface recognition). Information is communicated to the mainly dorsal interplay that starts *motor execution,* sending the output to the body. When the *mainly ventral interplay* doesn't recognize the *plastic effect*, as in the case of the *content depicted recognition,* the process of *action planning* stops. The output, in this case, is not sent to the body. The tension between these two simultaneous processes makes pictorial perception special for Ferretti. Both processes, however, are applications of the sandwich model of the mind. In both cases, there is an input (the picture), an internal process of computation (mainly ventral interplay and mainly dorsal interplay), and an output (the body, present in the case of surface recognition and absent in the case of content depicted recognition).

I want to oppose the sandwich model to Gallagher's (2021) notion of cognition as a meshed architecture that describes theatrical, choreographic, athletic, and artistic performances. Gallagher borrowed this notion from Christensen, Sutton and McIlwain (2016). The three authors originally conceived the idea of *Mesh*, later renamed by Gallagher (2021) meshed architecture, as a model to describe skilled behaviors. A popular position in the philosophical debate (Dreyfus, Dreyfus 1986) says

that skilled behaviors are entirely automatic in the expert, while cognitive control plays a role only in skill learning. On the contrary, Christensen, Sutton, and McIlwain endorse a more holistic view of skilled behavior, for which the exercise of skill implies both cognitive capacities and instinctual bodily reactions, even in an expert case. In Mesh, a pivotal cognitive role is played by situation awareness, which directly controls the parameters of action (*action set* and *action gist*) during skill execution. The neuroscientific basis of the model is in the *systems view* conception of cognitive and automatic control. *Systems view* questions the dualism between mindful cognitive control and a mindless set of bodily reactions in opposition to the more popular *dual-process* conception (Evans, Stanovich 2013). Instead, the dual-process and the sandwich model (of which the *dual-process* model seems the natural prosecution) neatly separate an automatic body from the conscious cognitive process.

I suggest that a model like the *meshed architecture* can also be applied fruitfully to picture perception. Like many other abilities described by meshed architecture, picture perception is also composed of many *component skills* that are hierarchically organized. It consists of a visuomotor interaction with the picture formed of perceptual processes, motor activations, and more abstract interpretative processes that can re-orient the interaction in problematic cases. A model like the meshed architecture aims to directly link the *body-schematic processes* (an intelligent and flexible interaction) with the higher processes of cognition. Gallagher does not suggest that the higher processes control the bodily processes in a top-down way; rather, the simplest of bodily processes are already intelligent thanks to an intrinsic control that permits the *meshed architecture* to work in a bottom-up way. Indeed "body-schematic processes are perfectly specific, adaptive and highly dynamical in order to facilitate movement aligned with particular situations and for specific intentions" (Gallagher 2021: 47). I argue that this model of cognition can be applied to pictorial perception with good results as an alternative to the *sandwich model* of the *DVAAP*. The *meshed architecture* presents a model in which perception, action and cognition are not separated but on a continuous scale that allows fast and flexible information flows.

Picture perception is a process of visual reenactment that is refined, attuned, and adjusted thanks to a meshed architecture model[3] composed

[3] As mentioned, the meshed architecture model was initially developed to describe practical skills, such as those of a footballer or a golfer. One of the cognitive components in Christensen, Sutton, and McIlwain's original model is situational

of various cognitive abilities, from the simple act of perception to sensorimotor abilities and higher cognition. The latter does not control the process from the top but is rather called for help from the bottom when the interaction finds difficulties in comprehending the picture.

3. *Pictorial experience and aesthetic anti-autonomism*

According to Dreon and Vara Sanchez (2022) an essential feature of enactivist aesthetics that distinguishes this approach from other naturalist approaches to the discipline is an anti-autonomist conception of the arts. Anti-autonomism criticizes every form of autonomist position in aesthetics. Autonomism sees art as a class of products separate from any other human activity. The most systematic criticism of aesthetic autonomism is contained in Gadamer (1960). Gadamer sees the origin of this philosophical position in the third critic of Immanuel Kant (1790) and in the exclusion operated by Kant of cognitive factors in the appreciation of "free beauty", that is, in the appreciation of natural beauty. An autonomist account doesn't mean only a separation between aesthetics and epistemology, religion, and ethics but also a separation between art and ordinary life. In *Art as experience* (1989), a pivotal work in American pragmatist philosophy, John Dewey has, for the first time, underlined the risk of aesthetic autonomism. In *Performance/art* (2021), an enactivist take on the performing arts, Shaun Gallagher draws on Dewey's anti-autonomist position, searching for the common elements between art and ordinary life but also emphasizing what differentiates them. For Gallagher and Dewey, the aesthetic is a quality contained in every experience, very often implicit and, in rare cases, explicit.

To argue for an anti-autonomist conception of picture perception means to see it as an experience in continuity with ordinary experience. Most of the various picture perception accounts suggested (Wollheim 1987; Lopes 2005; Nanay 2015) argue that picture perception is a two-fold experience in which the two folds (*recognitional* and *configurational fold*) are distinguishable on an ontological level. According to this conception,

awareness, a capacity not implied in picture perception. The reader should note that the model is used in a much simpler form in this paper. Higher cognition should be intended here as a form of cultural awareness that does not imply a focus on the actual context in which the act of perception is executed. This cultural awareness comprises a series of historical and artistic notions that are part of the beholder's general knowledge.

when the beholder is looking at a painting, he has two simultaneous experiences: the experience of what is depicted and the experience of the surface of the picture. This is somewhat different from how we normally perceive the world in ordinary life. When I look at a rose, I don't have an experience of the red color that is neatly distinguishable from my experience of the object as a rose. Perception is not an association of atomic entities but a fully integrated complex from the start, a gestalt as Merleau-Ponty (1945) has written. In this paper, I endorse the idea that picture perception is fully integrated from the beginning for the beholder, in line with what Hopkins (2010) has called unitary accounts of seeing-in, like that of Walton (1990). In this way, we can obtain a conception of picture perception that is closer to that of ordinary experience. Configurational and recognitional folds are only abstractions rather than actual psychological folds. Stating otherwise would imply committing what Dewey called a philosophical fallacy as explained by Dreon (2022). The philosophical fallacy occurs when we give an ontological status and precedence to parts of the experience that are instead the results of an analysis a posteriori.

There is also a second untold assumption of many accounts of picture perception that I want to question here. What is this assumption? The conception of picture perception as immediate and apprehensible with a quick look. Instead, experimental studies on picture perception, like those of Yarbus (1967), show us that picture perception is a temporal process. It can last only a few seconds or twenty minutes but always implies a temporal duration. This is because, as perceptual activity theory and active vision accounts state, perception is an active exploration of the environment. And a process of active exploration cannot be immediate. Even the simplest of paintings cannot be grasped with a single eye movement, but it always implies an active exploration, even for a few seconds.

In the last paragraph of this paper, I want to show how this enactivist pragmatist model of picture perception works, taking as an example the process of perception of a pencil sketch (1890) by Vincent Van Gogh, *Interior with five figures around a table* (fig. 1).

4. *Vision, movement and abstract cognition in pictorial interaction*

Suppose the beholder has carefully observed this picture in the gallery. Then he goes out of the gallery and into the open air. Closing his eyes, he

feels capable of clearly stating that the picture in question depicts a group of men who have just finished a meal. How can an enactivist pragmatist account explain the picture perception the beholder has experienced?

For one of the perceptual theories affiliated with the concept of reenactment, the perceptual activity theory (Hochberg 1968), perception is a continuous process of exploring and probing the environment around us. Reenactment itself is a form of exploration. Here, the exploration is only imagined, given that no sensory feedback from the world exists. In the case of picture perception, a borderline case between imagery and ordinary perception, the process of perception is still an exploration, but this time aided by the scaffolding of the picture. Yarbus (1967) used an eye-tracker to record eye movements from various experimental subjects involved in various tasks. Some of these recordings were made during the inspection of photographs. Several photographs chosen for these experiments depict faces. In *Eye movements and vision* (1967) we can see the results of these recordings in graphs (Yarbus 1967: 179-81). Each of these graphs represents the eye movements of a single experimental subject on a photograph depicting a face. Some areas are highlighted in each graph, while others are not. Furthermore, the graphs look all very similar. Evidence from these results shows that even in front of pictures, the eye is attracted by the most informative parts of the photograph. This means that vision's exploratory and probing nature is also proper for picture perception. Secondly, the results reveal that the eye is always attracted by the same informative areas in front of a particular content depicted. Indeed, in the experiment, each scan path had a similar trend. The graphs traced by the eye-tracker show that the experimental subjects mainly looked at the most meaningful parts of the faces depicted, like the eyes, the mouth, and the nose. These are also the parts we tend to look at more when perceiving a real human face. In particular, the mouth and the eyes are the more salient parts of the face for us because they can be vehicles of meaning. Thus, in perceiving the picture, the beholder is reenacting the same eye movements he would have enacted in front of a real human face. For Nanay and Ferretti this reenactment is simply the output of a representational process in the brain. Instead, I argue that the interaction between reenactments and pictures can explain most picture perception cases without involving mental representations of any kind.

Let's return to fig. 1. Suppose the beholder has not read the picture's title. He does not know whatsoever regarding the content depicted in the sketch. When perceiving Van Gogh's drawing, the beholder will look at the scene first, identifying the principal characters. This identification consists

of a test: in a second or a little more, the beholder reenacts the same perceptual schemas he usually enacts for detecting a drawn man. Past experiences with pictures have taught him simple schemas for recognizing a man drawn in a picture. Now that the drawn figure is identified, the beholder can, this time, reenact the same eye movements he would have enacted in front of a real man. He will look more at the characters' faces and arms, searching for visual expressions or possible action signals. Directing his eyes through attention to the same informational areas he is attracted to when looking at a real man, he gives imaginary reality to the men sketched by Van Gogh, *seeing* the pencil strokes *as* a group of men. Then, he will look at some signs of the pencil on the right under one of the figures. He will verify the presence of a chair by reenacting schemas aimed at recognizing drawn chairs. He will orient his attention to the chair's legs and back. Only three legs are present, and this absence of perceptual feedback could testify to the beholder that the lines drawn cannot be seen as a chair. But by reenacting the same eye movements he would have enacted in front of a real chair, he understands that the chair's fourth leg is beyond the sitting man's left leg. This is because when we reenact a specific behavior, we also recall a specific sensorimotor knowledge that helps explain cases of amodal perception very present in pictures.

This theoretical hypothesis draws upon the three enactivist and pragmatist assumptions discussed in the previous sections of the paper. The first assumption is that picture perception does not occur in the beholder's brain. Picture perception is as much external as it is internal. The reenactment consists of eye movements composed of perceptual and sensorimotor faculties. Both of them are faculties outside the brain. This process of reenactment is a form of active perception that does not passively photograph what is depicted in the painting but instead identifies the object in the picture thanks to optical movements directed by attention, the eyes, and previously learned schemas (learned both during previous picture perceptions or in real life). The second assumption says that cognitive processes do not follow the sandwich model of the mind, but they rather imply a holistic integration of our different faculties (perception, action, cognition, etc.). In the *DVAPP*, the dorsal stream (linked to motor programming) obeys the orders of the ventral stream (linked to action planning and response selection functions) without any connection to perception in a rigid schema of input, computational process, and output. In my hypothesis, perception (visual sensation) and action (eye movement) interact from the beginning

in a process of visuomotor reenactment independently from higher cognitive processes. The last assumption says that the pictorial experience is not special and autonomous but in continuity with ordinary experience. My account is consistent with this assumption. Picture perception is in continuity with ordinary perception because the reenactment process we put in action in front of a picture is strictly related to the perceptual processes of ordinary life.

At the beginning of the paper, I briefly said that sometimes, in doubt cases, higher cognition can nonetheless contribute to picture perception. This contribution does not imply the use of representations. The unit of cognition (body, environment, brain) should be seen as an adjustment in which the brain can sometimes be attuned to the other components and readjust the relation between the body and the environment.

What are the occasions when perception is ambiguous and generates doubts in the beholder?

Let's go back to fig. 1. Some elements in the painting are identifiable without ambiguity, and they can be easily recognized by the beholder (the human figures, their arms, their legs, and the table around which they sit together). Other elements are more obscure. Look at the circular pencil strokes on the left of the man with the hat. These lines can depict all equally possible things: a cloud, a comic strip balloon, and an airflow from the open window. This part of the painting is ambiguous. The beholder can solve the enigma only with the help of reasoning: the men around the table are finishing their meal. The circular pencil strokes represent a puff of smoke from the man with the hat. Very often, smokers light their cigarette or their pipe at the end of a meal. This hypothesis is verified through reasoning. On the contrary, the beholder does not resort to reason when recognizing the legs of the man with the hat or the teapot in the hands of the woman. We get, in fact, these things right just if we apply the proper *schema,* the right reenactment that orients our attention in the right way.

The recognition of the puff of smoke can be possible only thanks to a *meshed architecture* connecting flexible visuomotor responses at the bottom with higher cognition at the top and vice versa. After the beholder conceptually identifies the pencil strokes as a puff of smoke, the relation between the eye movements and the painting is readjusted. Now, the beholder can reenact the eye movements he usually put in action when recognizing a drawn puff of smoke. He will focus on certain features of the pencil strokes instead of others, shaping his perception according to the conceptual knowledge of higher cognition.

John Dewey, one of the authors upon which the actual pragmatic turn in cognitive science draws, has underlined in *Art as experience* (1934) how the aesthetic experience often includes moments of tension. At this point for Dewey reasoning comes into play. The experience flowing unaware of itself is now completely aware and uses reason and abstraction to overcome the obstacle. I have shown in this last paragraph that sometimes, in picture perception, we can find a similar situation. The beholder interacts with the painting through consolidated habits that enable him to recognize what is depicted. He does not use higher cognition, only his eyes and body. When the recognition process finds an ambiguous element in the painting, eye movements have found an obstacle on their path, an obstacle that interrupts the process of visuomotor interaction. Only higher cognition can help formulate and verify conceptual hypotheses in this case[4]. When this process of interpretation ends, the beholder returns to interact with the picture, attuning and refining the reenactment process.

Conclusions

I have suggested a new hypothesis concerning how picture perception works. This hypothesis is consistent with the central tenets of enactivism and pragmatism. I have also underlined some weaknesses of the *Dorsal ventral account of picture perception* of Ferretti (2018; 2021) and Nanay (2010; 2011; 2015). First, I have argued that picture perception is not an internal process. It consists of a visuomotor interaction with the painting that usually happens outside the head. This interaction is a visuomotor reenactment that helps the beholder *see* some blots on the canvas *as* a depiction of x. Secondly, I have suggested that picture perception is not a psychological process divided into three distinct parts (like the sandwich model of cognition suggests) but rather a holistic architecture that follows the structure of the *meshed architecture model*. This architecture comprises various component skills: perception, sensorimotor activations, and higher cognition. In the third stance, I have suggested, in

[4] It can be argued that this kind of conceptual knowledge necessarily implies representations. Still, it should be noted that the knowledge I am referring to is not *decoupled* from the body. It is recalled by the brain only with the practical goal of refining the interactive process. Thus, it should not be conceived as a form of *off-line* cognition.

line with aesthetic anti-autonomism, that picture perception is an aesthetic experience in continuity with ordinary experience. Indeed, to reenact specific eye movements means to repeat a particular visual schema enacted in ordinary perception without the original perceptual feedback. In the paper's last part, I analyzed a concrete case of picture perception: the visual perception of a sketch by Vincent Van Gogh, *Interior with five figures around a table*. In inspecting the content depicted, the beholder reenacts specific visual schemas that detect relevant information in the picture. These visual schemas are similar to those enacted in ordinary life. Indeed, the beholder searches in the picture for the same visual information he would have searched for in front of the real visual scene. Sometimes, the scan path can end abruptly due to visual ambiguities. Thanks to the meshed architecture, higher cognition is called for help from the bottom. By formulating different hypotheses, higher cognition readjusts the relation between eye movements and the painting. Now, the process of interaction can restart.

Bibliography

Adams, F., Aizawa, K., *The bounds of cognition,* Hoboken NJ, Blackwell Publishing, 2008.

Bajcsy, R., *Active perception,* "Proceedings of the IEEE", 76 (1988), pp. 996-1005.

Bedia, M.G., Aguilera, M., Gomez, T., Larrode, D.G., Seron, F., *Quantifying long-range correlations and 1/f patterns in a minimal experiment of social interaction,* "Frontiers in psychology", 5 (2014), 1281.

Chomsky, N., *Language and nature,* "Mind", 104/413 (1995), pp. 1-61.

Christensen, W.D., Sutton, J., McIlwain, D.J.F., *Cognition in skilled action, meshed control and the varieties of skilled experience,* "Mind & language", 31/1 (2016), pp. 37-66.

Dewey, J., *The reflex-arc concept in psychology* (1896), in J.A. Boydston (ed.), *The early works*, vol.5*,* Carbondale and Edwardsville, Southern Illinois University Press, 1972, pp. 96-109.

Dewey, J., *Art as experience* (1934), in J.A. Boydston (ed.), *The later works*, vol. 10*,* Carbondale and Edwardsville, Southern Illinois University Press, 1989.

Di Paolo, E., Burhmann, T., Barandiaran, X., *Sensorimotor life: an enactive proposal,* Oxford, Oxford University Press, 2017.

Dotov, D.G., Nie, L., Chemero, A., *A demonstration of the transition of ready-to-hand to unready-to-hand,* "PLoS ONE", 5 (2010), e9433.

Dreon, R., *Human landscapes: contributions to a pragmatist anthropology,* New York, State University of New York Press, 2022.

Dreon, R., Vara Sanchez, C., *Naturalist trends in current aesthetics,* "Studi di estetica", 22/3(2022), pp. 221-44.

Dreyfus, H.L., Dreyfus, S.E., *Mind over machine: the power of human intuition and expertise in the era of the computer,* New York, Free Press, 1986.

Evans, J. St. B.T., Stanovich, K.E., *Dual-process theories of higher cognition: advancing the debate,* "Perspectives on psychological science", 8 (2013), 223-41.

Ferretti, G., *Through the forest of motor representations,* "Consciousness and cognition", 43 (2016b), pp. 177-96.

Ferretti, G., *Visual feeling of presence,* "Pacific philosophical quarterly", 99/S1 (2016c), pp. 112-36.

Ferretti, G., *The neural dynamics of seeing-in*, "Erkenntnis", 84 (2019), pp. 1285-324.

Fodor, J., *Methodological solipsism considered as a research strategy in cognitive science,* "Behavioral and brain sciences", 3 (1980), pp. 63-73.

Foglia, L., O'Regan, K., *A new imagery debate: enactive and sensorimotor accounts,* "Review of philosophy and psychology", 7/1 (2016), pp. 181-196.

Gadamer, H.-G., *Truth and method* (1960), Engl. transl. W. Glen-Doepel, revised by J. Weinsheimer, D.G. Marshall, London, Bloomsbury Academic, 2013.

Gallagher, S., *How the body shapes the mind*, Oxford, Clarendon Press, 2005.

Gallagher, S., *Enactivist interventions: rethinking the mind*, Oxford, Oxford University Press, 2017.

Gallagher, S., *Performance/art. The venetian lectures*, Sesto San Giovanni, Mimesis international, 2021.

Gallagher, S., Zahavi, D., *The phenomenological mind* (2007), London, Routledge, 2021.

Gibson, J., *The ecological approach to visual perception* (1979), New York, Taylor & Francis, 2014.

Goodale, M., Milner, A.D., *Sight unseen*, Oxford, Oxford University Press, 2004.

Harman, G., *Wide functionalism,* in S.R. Schiffer, S. Steele (eds.), *Cognition and representation,* Boulder (Colorado), Westview Press, 1988, pp. 11-20.

Hochberg, J., *In the mind's eye,* in R.N. Haber (ed.), *Contemporary theory and research in visual perception,* New York, Holt Rinehart and Winston, 1968, pp. 309-31.

Hopkins, R., *Inflected pictorial experience its treatment and significance,* in C. Abell, K. Bantinaki (eds.), *Philosophical perspectives on depiction,* Oxford, Oxford University Press, 2010, pp. 151-80.

Hurley, S., *Consciousness in action* (1998), Cambridge (MA), Harvard University Press, 2002.

Hutto, D.D., Myin, E., *Radicalizing enactivism: basic minds without content,* Cambridge (MA), MIT press, 2012.

Johansson, R., Holsanova, J., Holmqvist, K., *Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and complete darkness,* "Cognitive science", 30 (2006), pp. 1053-79.

Kant, I., *Critique of the power of judgment* (1790), ed. by P. Guyer, Engl. transl. P. Guyer, E. Matthews, Cambridge, Cambridge University Press, 2000.

Kelso, J.A.S., Dumas, G., Tognoli, E., *Outline of a general theory of behavior and brain coordination,* "Neural networks", 37 (2013), pp. 120-31.

Kirchhoff, M.D., *Extended cognition and the causal-constitutive fallacy: in search for a diachronic and dynamical conception of constitution,* "Philosophy and phenomenological research", 90/2 (2015), pp. 320-60.

Kosslyn, S.M., Thompson, W.L., Kim, I.J., Alpert, N.M., *Topographical representations of mental images in primary visual cortex,* "Nature", 378 (1995), pp. 496-8.

Lopes, D.M., *Sight and sensibility,* Oxford, Oxford University Press, 2005.

Merleau-Ponty, M., *Phénoménologie de la perception*, Paris, Gallimard, 1945.

Milner, A.D., Goodale, M., *The visual brain in action* (1995), Oxford, Oxford University Press, 2006.

Milner, A.D., Goodale, M., *Two visual systems re-viewed,* "Neuropsychologia", 46 (2008), pp. 774-85.

Nanay, B., *Inflected and uninflected experience of pictures,* in C. Abell, K. Bantinaki (eds.), *Philosophical perspectives on depiction,* Oxford, Oxford University Press, 2010, pp. 181-207.

Nanay, B., *Perceiving pictures,* "Phenomenology and cognitive sciences", 10 (2011), pp. 461-80.

Nanay, B., Trompe l'oeil *and the dorsal/ventral account of picture perception,* "Review of philosophy and psychology", 6 (2015), pp. 181-97.

Noë, A., *Is this world a grand illusion?*, "Journal of consciousness studies", 9/5-6 (2002), pp. 1-12.

Noë, A., O'Regan, K., *A sensorimotor account of vision and visual consciousness,* "Behavioral and brain sciences", 24/5 (2001), pp. 883-917.

Perry, C.J., Fallah, M., *Feature integration and object representation along the dorsal stream visual hierarchy,* "Frontiers in computational neuroscience", 8 (2014), article 84 (https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2014.00084/full; accessed 03.2025).

Riley, M.A., Richardson, M., Shockley, K., Ramenzoni, V.C., *Interpersonal synergies,* "Frontiers in psychology", 2 (2011), 38 (https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2011.00038/full; accessed 03.2025).

Shockley, K., Santana, M.-V., Fowler, C.A., *Mutual interpersonal postural constraints are involved in cooperative conversation,* "Journal of experimental psychology: human perception and performance", 29 (2003), pp. 326-32.

Shusterman, R., *Pragmatist aesthetics. Living beauty, rethinking art* (1992), Lanham etc., Rowman and Littlefield Publishers, 2000.

Thomas, N.J.T., *Are theories of imagery theories of imagination? An active perception approach to conscious mental content,* "Cognitive science", 23 (1999), pp. 207-45.

Thomas, N.J., *Imagery and the coherence of imagination,* "Journal of philosophical research", 22 (2010), pp. 95-127.

Thompson, E., *Mind in life: biology, phenomenology and the sciences of life,* Cambridge (MA), Harvard University Press, 2007.

Van Orden, G.C., Holden, J.G., Turvey, M.T., *Self-organization of cognitive performance,* "Journal of experimental psychology", 132 (2003), pp. 331-50.

Varela, F., Thompson, E., Rosch, E., *The embodied mind: cognitive science and human experience* (1991), Cambridge (MA), MIT press, 2017.

Verhoef, B.E., Vogels, R., Janssen, P., *Synchronization between the end stages of the dorsal and the ventral visual stream,* "Journal of neurophysiology", 105/5 (2011), pp. 2030-42.

Vishwanath, D., *Information in surface and depth perception: Reconciling pictures and reality*, in L. Albertazzi, G.J. Van Tonder, D. Vishwanath (eds.), *Perception beyond inference. The information content of visual processes,* Cambridge (MA), MIT press, 2011, pp. 201-40.

Von Uexküll, J., *A foray into the worlds of animals and humans* (1934), in *A foray into the worlds of animals and humans: with a theory of meaning*, Engl. transl. J.D. O'Neil, Minneapolis, University of Minnesota Press, 2010.

Walton, K.L., *Mimesis as make-believe: on the foundations of the representational arts*, Cambridge (MA), Harvard University Press, 1990.

Westwood, D. *et al.*, *Grasping two-dimensional images and three-dimensional objects in visual-form agnosia,* "Experimental Brain Research", 144/2 (2002), pp. 262-7.

Wollheim, R., *Painting as an art* (1987), Princeton, Princeton University Press, 2023.

Yarbus, A., *Eye movements and vision,* New York, Plenum Press, 1967.

Zipoli Caiani, S., Ferretti, G., *Semantic and pragmatic integration in vision for action,* "Consciousness and cognition", 48 (2016), pp. 40-54.

Fig. 1. V. Van Gogh, *Interior with five figures around a table* (1890), pencil on paper, 23.2x32cm, Van Gogh Museum, Amsterdam.