

ARTICLE OPEN ACCESS

The Dorsal-Ventral Account of Picture Perception

Gabriele Ferretti^{1,2} 

¹Università di Bergamo, Bergamo, Italy | ²Institut für Philosophie II, Ruhr-University Bochum, Bochum, Germany

Correspondence: Gabriele Ferretti (fairg@live.it; gabriele.ferretti@unibg.it)

Received: 16 December 2025 | **Revised:** 16 December 2025 | **Accepted:** 26 February 2026

ABSTRACT

What is the nature of our perception of pictures? Philosophers intrigued by this question, and adopting a naturalistic perspective, have turned to findings from visual neuroscience to answer it. This perspective seeks to address the question within the framework of the *Two Visual Systems Model*, which provides a specific anatomo-functional description of how our visual system operates. According to this model, the visual cortex hosts a specific hodological division between a ventral stream, responsible for visual recognition, and a dorsal stream, responsible for the visual guidance of action. A philosophical account of picture perception based on this model is therefore referred to as the *dorsal/ventral account of picture perception*. Since its inception, this account has evolved in parallel with the advancements in the *Two Visual Systems Model* itself. This paper surveys the various interpretations under the different formulations of this account and traces their development within the literature of empirically informed philosophy of cognitive science.

1 | Introduction: A New Approach to an Old Problem

What is the nature of our perception of pictures? Or, as Bence Nanay puts it: “What perceptual state are we in when we see an object in a picture?” (2011, 461). If you think science can help philosophers in developing their theories, then you are prone to look at the biological functioning of the visual system to flag the possible answers to this question. From Nanay (2011) himself comes the explicit and full proposal of considering this question under a naturalistic stance.

If by endorsing a naturalistic stance we mean to refer to science in a general sense, then aesthetics and art have always been steeped in scientific knowledge about the functioning of the visual system, as well as in optics and psychophysics. However, since 2005, this means referring to one of the most impactful and still debated models in visual neuroscience, the *Two Visual Systems Model* (henceforth, TVSM), developed by Milner and Goodale (1995/2006), based on the original insights by Ungerleider and Mishkin (1982), suggesting the anatomo-functional division, within our visual system, between a ventral and a dorsal stream. The philosophical account of picture perception

that follows the TVSM is called, not by chance, the *dorsal/ventral account of picture perception* (henceforth, DVAPP) (Nanay 2015).

I am saying (2005) because, as Nanay admits, “it was Matthen who first suggested the possibility that the ventral/dorsal distinction may be relevant for understanding what happens if we perceive pictures—and I owe the basic idea of this paper to him” (477). Matthen (2005) has indeed been the first to propose to embrace this model to understand something more about picture perception.

The merit of Nanay, who tried to bring to fruition what Matthen had already started, was then twofold. First, explicitly offering an account of how the two streams work in front of a picture, refining the proposal by Matthen. Second, unifying the TVSM with the philosophical intuitions of one of the most famous theories in the philosophy of the perception of pictures, the theory of *seeing-in* and *two-foldness* proposed by Wollheim (1998).

However, the TVSM has massively evolved through the years, proposing different sets of evidence on the functioning of each

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Philosophy Compass* published by John Wiley & Sons Ltd.

stream, as well as on their relation. Bringing together the important insights from philosophers into a coherent view of picture perception, in the light of the most up to date TVSM, was the task of more recent formulations of the DVAPP (Ferretti 2016a, 2018b), which tried to extend Nanay's and Matthen's proposals with new insights on the streams, while being faithful to the Wollheimian tradition.

The DVAPP has the merit of combining the philosophical speculation on picture perception with our current knowledge of the neurobiology of the visual system. This article discusses the evolution of this account, which follows the evolution of the TVSM.

2 | Seeing-In and Two-Foldness

The DVAPP characterizes the perceptual state we are in during picture perception by following the notion of *seeing-in*, related to *two-foldness*, offered by Wollheim (1998). Picture perception hosts two folds: a *configurational fold*, that is, the surface (vehicle) of the picture, and a *recognitional fold*, that is, the content in the pictorial space. This two-foldness leads us to see something in a picture, that is, seeing-in.

According to this theory, we simultaneously see both the picture's surface and the pictorial space visually encoded across it. However, although Wollheim talked about “simultaneous awareness” (see Nanay 2011, 462, 463; on Wollheim 1980, 214–215), DVAPP's simultaneity is about representation, not awareness. The shift from awareness to representation offers some crucial theoretical advantages (Nanay 2011, 462, 463; 2015, 192; 2017). First, *seeing-in* does not entail that we are always aware¹ of both the folds, simultaneously, but simply that our visual system consciously or unconsciously represents both of them, simultaneously. This overcomes the worry that we cannot consciously perceive, simultaneously, some properties of the surface and some of the depicted object.² This notion also makes justice to the fact that, most of the time, we are not aware of the surface, when appreciating a pictorial content. Thus, this notion grants two-foldness, permitting seeing-in, via simultaneous representation, while avoiding problems with simultaneous consciousness. Finally, as suggested in a later formulation of the DVAPP (Ferretti 2018b; Section 8), maintaining simultaneous representation, even without simultaneous consciousness, is crucial: we need to visually represent the surface in order to perceive a pictorial space as such; indeed, when we cannot visually track the presence of a surface, we fall into the illusion of being in front of an object in the flesh, as happens with pictorial illusions such as *trompe l'oeils*.³

The crucial contribution of the DVAPP has been the attempt to naturalize, in the light of our current knowledge on the neurobiology of the visual system, the notion of *seeing-in*, related to *two-foldness*. This means explaining the peculiarity of our perception of pictures by looking at what's happening within our visual brain, considering the evidence coming from the most important model within vision neuroscience, the TVSM. At this point, to appreciate the DVAPP, we must first analyze the TVSM.

3 | The Two Visual Systems Model

The TVSM has a long history of theorization based on the evidence found by different generations of neuroscientists. Long story made short, we can conceive two main formulations of this model. One is the *segregation view*, and the other is the *integration view*. Analyzing these two formulations will help to understand the evolution of the DVAPP in parallel with the advancements of the TVSM itself.

According to the first TVSM (segregation view), there are two segregated visual pathways in our visual cortex (Milner and Goodale 1995/2006). On the anatomical level, there is a ventral stream, that is, the occipito-temporal network, which goes from the primary visual cortex to the infero-temporal cortex, and a dorsal stream, that is, the occipito-parietal network, which goes from the primary visual cortex to the posterior parietal cortex, and exhibits specific connections to the pre-motor areas. On the functional level, the ventral stream allows primarily conscious, but also unconscious, visual object recognition, which subserves perception from an allocentric frame of reference. Conversely, the dorsal stream allows the (uniquely) unconscious visual guidance of action, and the related attribution of action properties to the objects we perceive, which subserves perception from an egocentric frame of reference, within the peripersonal space (the space of action) of the observer.

This initial version of the TVSM has been reconsidered based on new sets of evidence (integration view). It has been shown, at the anatomical level, that (i) the streams have anatomical connections; (ii) the dorsal stream can be further divided into a dorsal portion (the dorso-dorsal stream) and a ventral portion (the ventro-dorsal stream), and the latter offers dorsal responses mediated by a shade of ventral processing. This aligns to a new description of the functional level: (iii) many tasks in both visual recognition and visual guidance of action are subserved by an interplay between the streams. Accordingly, (iv) the ventral stream, per se, can be involved in action, whereas the dorsal stream, per se, can be involved in recognition (Sakreida et al. 2016; Rossetti et al. 2017; Ferretti 2018b).

We'll get back to technical details about the streams later. Now, DVAPP's first formulations endorse functional dissociation between the streams, following the initial TVSM, whereas the last DVAPP acknowledges an interplay between them, following the last TVSM.

4 | The Dorsal/Ventral Account of Picture Perception, Ante Litteram

As Nanay himself noted (Section 1), Matthen (2005, see also 2010) has been the first to conjugate the TVSM with the philosophical reflection on pictures. Matthen invokes the TVSM but decides to use a general functional label, related to *motion-guiding-vision* versus *descriptive vision*, rather than a neurophysiological one, related to *dorsal* versus *ventral* visual processing. This permits to focus on visual functions in a more abstract manner without (and independently of the) reference to

the neurobiological geography of the neural loci within which these functions are realized (2005, 297).

The main claim of this framework is that *motion-guiding-vision* is attuned to the surface, whereas *descriptive vision* is attuned to the depicted object. “Pictures evoke in us an awareness of the same descriptive visual features of objects as real-life scenes—features that we use for epistemic purposes. From a feature-attribution point of view, many things that can be seen in real life can also be seen in a picture, and in the same way: When we look at a picture, this kind of input accounts for why we construct a three-dimensional representation of the depicted object side-by-side with one of the picture itself. What then is the phenomenological difference between a “hyper-realist”⁴ picture and real life? The standard story appeals to the flatness cues that pictures provide. However, the engagement of motion-guiding vision is another important difference. We have just seen that though we are able accurately to report on the visual features of depicted objects, we are unable to make appropriate bodily motions with respect to them. This difference entails that we do not have agent-centered experiences of depicted things. When we see a thing in a picture, we cannot visually locate it in ways that guide bodily motion with respect to it” (Matthen 2005, 313).

Although pictorial objects solicit a recognition that is similar to objects in the flesh, a big difference with concrete objects is that pictorial objects do not allow us to perceive them as available for action, and indeed they are not available for action (cf. below Section 6 on this distinction). The investigation on picture perception is an investigation on visual space. We cannot locate pictorial objects with respect to ourselves (i.e., our body), and so with respect to our actions, and our view of them does not change in relation to our movement: only surfaces offer these spatial aspects (Matthen 2005, 313–319). Thus, depicted objects cannot have any significance for motion-guiding vision but only for descriptive vision. Motion-guiding vision can represent only the surface.

Nanay’s DVAPP starts from these insights but offers a more technical description.

5 | The Dorsal/Ventral Account of Picture Perception

The DVAPP is the TVSM applied to picture perception. Then, it must account for the behavior of the visual streams’ processing in relation to the notion of *seeing-in* based on *two-foldness à la Wollheim*. This means considering that (a) our visual system is divided into two streams, and that (b) picture perception hosts two main perceptual components, occurring simultaneously, that is, surface perception and perception of the pictorial space. Thus, the main task for this account is to explain the relation between each stream’s processing with respect to each of these two components of picture perception.

In his paper manifesto of this account (2011), Nanay explicitly inaugurates the DVAPP by developing these insights, leading Matthen’s germinal intuition to blossom. He already suggested

that “any account of seeing-in must be able to tell how this experience represents the picture’s surface and how it represents the depicted object” (Nanay 2010a, 199; see Hopkins 1998 on this necessity). In this first paper on the DVAPP, Nanay suggests that looking at the TVSM clearly shows that “the twofold experience of pictures Wollheim talks about corresponds to the dichotomy between our dorsal visual processing of the surface of the picture and our ventral visual processing of the depicted scene” (Nanay 2011, 464).

With respect to the formulation of the TVSM, Nanay (2011, 465) assumes functional difference between the streams while remaining neutral on potential anatomical segregations: there is a visual recognition subserved by the ventral stream and a vision-for-action subserved by a dorsal stream—this may be seen as compatible with the *segregation hypothesis* about the functions of the streams (Section 3).⁵

In this framework, Nanay offers four claims at the basis of the DVAPP, for which he discusses empirically based arguments that aim at *explaining* picture perception, starting by *describing* it in terms of ventral and dorsal processing being related to the perception of the surface and of the depicted object.

- a. The depicted scene is represented by ventral perception.
- b. The depicted scene is not represented by dorsal perception.
- c. The picture’s surface is represented by dorsal perception.
- d. The picture’s surface is not necessarily represented by ventral perception.

Seeing-in is based on the simultaneous visual representation of the surface and the depicted object. Ventral processing is responsible for visual object recognition. Then, it must be attuned onto the pictorial space, that is, (a). The surface can be represented by ventral perception. However, most of the time, we do not look at the surface, ignoring it, that is, (d). But the surface is an object we can act upon. Then, dorsal processing, responsible for vision-for-action, represents the surface, its action properties, and its spatial relation with respect to us, that is, (c). Finally, as we cannot interact with the pictorial space, and it is not possible to represent it, spatially, with respect to us, dorsal perception cannot represent it, as the pictorial space is not the space of action (there is no action property), and does not offer any spatial reference with respect to us (on which dorsal processing operates), that is, (b).

Thus, Nanay builds an account of picture perception in terms of dorsal and ventral vision, with reference to their functions, on the surface and the depicted scene. From this four-party claim, it follows that TVSM-based seeing-in, via two-foldness, is realized in two scenarios (Nanay 2011, 466, 467):

1. Dorsal vision represents the surface, whereas ventral vision represents the depicted object.
2. Ventral vision represents both the surface and the depicted object (the case of *inflection*,⁶ cf. Sections 6 and 7).

These conditions grant the *simultaneity* needed for *seeing-in*, with *two-foldness* maintained by at least one visual pathway

processing the depicted object and one visual pathway (not necessarily the same one) processing the picture's surface. This, with the proviso that although ventral processing can perform recognition on both the surface and the pictorial space, *as per* (a) and (d), the vision-for-action performed by dorsal processing can be exercised only on the surface, *as per* (b) and (c).

The DVAPP, then, offers a preliminary difference between picture perception and face-to-face perception. In front of an object in the flesh, both of our visual pathways, the ventral and the dorsal, are processing the whole object in relation to visual recognition and visually guided action. In front of a picture, our visual pathways behave differently with respect to the different components of the picture. The most important difference being that dorsal vision-for-action cannot represent the pictorial space.

Nanay (2011, 477) also discusses the differences between his account and Matthen's proposal. The most relevant are the following.

First, in relation to (c), Matthen suggests that the surface *can* be visually represented by motion-guiding vision (which translates, in Nanay's jargon, into dorsal processing). Nanay suggests that the surface *must* be visually represented by dorsal processing. This is because the spatial relation of the observer with respect to the surface is different compared to the spatial relation of the observer with respect to the pictorial object: while we move, we can appreciate perspectival changes in shape with respect to the surface, but not to the pictorial space. For Nanay, this is explained by invoking a perception of the surface, with respect to us, compensating for the view of the pictorial space, and this perception is courtesy of dorsal vision. As dorsal processing deals with action, and is responsible for the encoding of spatial relations between us and the perceived object, it has a special connection to space processing with respect to surfaces: It represents them for action, and thus their spatial relation with respect to us (this compensating for the pictorial space; which, indeed, cannot be represented dorsally) (Nanay 2011; cf. the end of Section 7).

Second, in relation to (b), for Matthen lack of motion-guiding vision onto the depicted object is mainly because we cannot interact with it. For Nanay, the argument can be sharpened. We cannot locate ourselves with respect to the pictorial space, that is, we cannot scale the distance between us and the pictorial space, and moving with respect to it will not bring into view any egocentric spatial information of this kind. Thus, dorsal processing cannot represent the depicted object and its spatial relation (also useful for action) with respect to us, that is, (b)—in accordance with the compensation described concerning (c).

These two points explain the nature of the spatial encoding of dorsal processing, suggesting why it *must* be attuned onto the surface, while it can not be attuned onto the pictorial space.

Third, though Matthen was the first to flag the possibility of a TVSM-based analysis of picture perception (Section 1), he was not interested in giving a global account of picture perception based on this model, as Nanay, but only in claiming that

depicted objects are not represented by motion-guiding vision (Nanay 2011, 477).

Finally, although Nanay's (2011) represents the manifesto of the DVAPP, this name (i.e., DVAPP) officially enters the literature with his work from (2015). In this paper, Nanay offers a specification concerning his (2011) while also suggesting the interdisciplinary benefits of building a theory of picture perception following the TVSM. The specification concerns possible misunderstandings on (b). Matthen (2005) suggested (in his jargon) that (b) is necessary. Although in his (2011, 468–469) Nanay seems to suggest that (b) is always the case, in his (2015, 189–190) he specifies that (b) may not be the case in some circumstances: The depicted scene could be represented by dorsal perception. In the special case of *trompe l'oeils*, which offer the illusory impression of being in front of an object in the flesh that offers interaction, (b) is violated: What explains the illusion is precisely that the depicted scene is represented by dorsal perception. Indeed, this paper focuses on the DVAPP's benefits in expalining *trompe l'oeils*.⁷

Thus, all in all, dorsal processing subserving vision-for-action guidance seems to be a solid ingredient to explain the differences between *face-to-face* and picture perception. And this also explains the differences, concerning the equilibrium of seeing-in, between standard pictures and illusory pictures *à la trompe l'oeil*. This, within the Wollheimian tradition upgraded with the TVSM.

6 | A First, Empirical and Explanatory Extension of the Account

The DVAPP is built on TVSM-based assumptions on the way our visual system works in front of a picture (differently from a concrete object). Since ventral processing performs visual recognition, it represents the depicted object and can also represent the surface, for in our perception of pictures we see (i.e., we visually represent) pictorial contents and their surface vehicles. Since dorsal processing performs visually guided action, it represents the surface, but not the depicted object, for in our perception of pictures we can motorically interact with the material surface we can scale our distance and gain perspective with respect to, but not with a pictorial space, as this is not materially localized within a space for action specifying spatial coordinates with respect to us.

An interesting fact, in this respect, is that most of the stimuli involved in neuroscience are pictorial stimuli. Given the difference, mentioned above, between perceiving an object in a picture and perceiving an object in the flesh, one may wonder about how licit this experimental methodology is (Ferretti 2017b; Snow et al. 2011): for example, the aim of visual neuroscience is that of recollecting data on how the visual system represents the external world, but the tests, rather than being based (only) on ecological stimuli, are (mostly) based on pictures. One line of answer may suggest that, except for action, it is safe to say that pictures host most, even if not all, of the optical and visual aspects of the external world, and so they are

reliable stimuli for understanding visual brain functions, especially when the investigation also recruits some ecological conditions (Ferretti 2017b).

Now, the methodology, in neuroscience, of using pictorial stimuli when testing perception has a twofold impact on the DVAPP.

It has a benefit for (a), as most of ventral processing is tested upon depicted stimuli in the lab (Snow et al. 2011). This confirms the activity of ventral processing at the basis of the recognition of a pictorial space and aligns to common sense: We recognize depicted objects, and they preserve most of the visual properties of objects in the flesh.⁸

However, as pictorial stimuli can tell something on the behavior of our visual system that is reflected on objects in the flesh, and since there are experiments on ecological conditions, (d) seems to be safe as well: ventral processing may, or may not, be involved in the perception of the surface (i.e., a concrete object). This also aligns to common sense: we recognize surfaces (and other concrete objects).

Furthermore, there is evidence that, when ventral processing is broken (causing *visual agnosia*), we fail to recognize both concrete and pictorial objects (Farah 2004; Jacob and Jeanerod 2003; cf. also Nanay 2011, 2015). This confirms the two claims above.

On the same direction, there is massive evidence that, when dorsal processing is broken (causing *optic ataxia*), visually guided action performed toward objects in the flesh is impaired. This seems to align to (c), as we can visually guide our action toward the surface, which is a concrete object, and so vision-for-action is attuned onto it, encoding our spatial relation with it.

However, the methodology of using pictures in neuroscience also offers food for thought to the DVAPP, when it comes to (b). Indeed, there are many sets of evidence on dorsal responses, related to action, in case of pictorial objects (non-illusory, i.e., non *trompe l'oeil*, usual depicted objects): in most cases, also the visual stimuli used to test action responses in the lab are pictorial (Ferretti 2016a, 2017b). These responses indicate that the dorsal stream can trigger motor preparation, related to the detection of action properties, also when the perceiver is in front of a depicted object.

This highlights a particular feature of the dorsal stream: It does not respond only to objects suitable for overt action execution, but it is also involved in motor preparation related to covert action, which can be triggered by different kinds of objects, independently of them being pictorial or concrete.

This body of evidence⁹ on dorsal responses to pictorial stimuli represents an opportunity for extension, for the DVAPP, considering the neuroscience of vision.

With this aim, Ferretti (2016a) offers a review on the many sets of evidence showing that the visuomotor circuits for grasping hosted within the dorsal stream (more precisely, within the AIP-F5 parieto-premotor circuit, related to the ventro-dorsal stream,

i.e., the ventral part of the dorsal stream, cf. Section 3) can respond to pictorial shapes, triggering the representation of the motor act that could be used, on action properties, if those shapes were concrete.¹⁰ Of course, it is shown that the motor acts, related to action properties, dorsally represented are not those recalled by a surface (e.g., a precision grip with two fingers) but by the object in the pictorial space (e.g., a power grip, on a cup, with the whole hand).

This leads to extend DVAPP's commitment to (b): Depicted objects *can* and indeed *are* represented by dorsal processing.

b*. The depicted object is represented by dorsal perception.

This move, however, begs a crucial question. How do we distinguish between pictures and concrete objects, then?¹¹ Let us explore the different facets of this point.

Nanay's version of the DVAPP suggested that having dorsal vision *only* onto the surface (i.e., not on the depicted object, i.e., b), that is, (c), and having ventral vision onto the depicted object, that is, (a), could grant a TVSM-based two-foldness for successful seeing-in. And this is what permits to discriminate between these two classes of perceptual objects: With objects in the flesh, the object we recognize is the same we can interact with, differently from depicted objects, which we recognize, but which we cannot interact with. Thus, although a concrete apple needs both stream's processing on this unique object, with a picture of an apple, dorsal processing represents the surface, but not the depicted object, which is ventrally represented.

However, the evidence on dorsal responses to depicted objects suggests an extension of the DVAPP (Ferretti 2016a). Recognition devoted to establishing what is the surface and what is the depicted object (as well as, on top of it, to distinguishing between objects in the flesh and pictures of objects), for interactivity, is not courtesy of dorsal processing. It is ventral processing that, with its sharper computational mechanisms devoted to object recognition, permits to distinguish between non-interactable pictorial spaces and surfaces offering interaction (i.e., the notion of *response selection of ventral processing*). Dorsal processing does not have the computational resources for this task. Plus, dorsal responses are faster than ventral ones. When exposed to a visual stimulus, the representation of a motor act that could be used with it is activated quite before the object is recognized as being 3-D or 2-D. Indeed, dorsal visuo-motor responses quickly occur, and then quickly decay (being inhibited), if we do not perform overt action. Accordingly, the target for action is recognized and selected by ventral processing, on whose identification dorsal motor response's actualization is conditional. This counts for both surfaces and pictorial spaces, as well as for objects in the flesh (Ibid.).

This specification dispels a potential doubt. Why is that, if dorsal processing responds to pictorial objects, also in case of standard, non-illusory pictures, we do not have the impression we can interact with them? This is because ventral processing is responsible for this impression, which depends on object reconstruction in recognition. Ventral processing distinguishes between pictorial objects and objects in the flesh and then flags the target onto which dorsal motoric responses, already

prepared, could be exercised. That is, dorsal responses are automatically triggered, independently of the stimulus and, without the ventral “pass,” they are not exercised: they are triggered and then inhibited to decay. This explains why we do not tend to interact with the pictorial space in the first place.

There is an important clarification here. One thing is to say that (O) we cannot interact with an object, this holding for an object in picture. Another is that (OO) we cannot perceive any possibility of interaction with an object, this holding for an object in picture. Although (O) appeals to overt action execution related to actual action performance, (OO) appeals to the perception of possibility of interaction, not necessarily followed by an action performance (Ferretti 2025). With standard pictures, both (O) and (OO) seem to be ruled out. With *trompe l'oeils*, however, (OO) is ruled out, even if (O) still holds: We have the impression we can interact with the object, even if it is a picture; precisely because it is an illusory picture. However, there is a third aspect of action (Jeannerod 2006; Jacob and Jeannerod 2003; Ferretti and Zipoli Caiani 2024): (OOO) our visuomotor system can elicit motor preparation (at the subpersonal, thus unconscious level) triggering the simulation of the motor response we could deploy with the shape we are in front of, were we in the position to act (in the case of pictures, were these pictorial objects concrete objects).¹² That is, it represents action properties. And this is what (b*) is about. Dorsal processing can trigger the motoric responses that could be used were the object happen to be a concrete object, effectively displaying the action properties recalled by it. Therefore, although we cannot interact with depicted objects, as per (O), and they do not foster any impression of interaction, as per (OO), this does not prevent motor preparation to take place, as per (OOO). This allows dorsal processing to represent the depicted object. Dorsal processing does not (strictly) depend on (O), and (OO), but is responsible for (OOO). Whereas (O) is a merely physical fact, it is ventral processing the one performing the visual process in (OO), as this is based on recognition, and dorsal responses *à la* (OOO) are always there. More precisely, (OO) is about perceiving the object in front of us as offering the effective possibility for interaction, that is, that it looks like a three-dimensional, solid, material, and concrete object we can interact with (a form of high-order recognition for effective interactability). Instead, (OOO) is about perceiving the specific action properties recalled by the object, for example, a power grip, or a precision grip, recalled by a handle (a form of visuomotor resonance on spatial properties). Ventral processing establishes whether we (perceive we) can interact with the object, as per (OO). Dorsal processing encodes the action properties of the object, as per (OOO).¹³

This extension, following (b*), makes the DVAPP more in tune with neuroscience, changing its implications, concerning the distinction between pictures and objects in the flesh. As for the perception of objects in the flesh, also with pictures both the streams can be involved, in the sense that both visual folds, the surface and the depicted object, are represented by both streams (differently from Nanay's account). The difference between these two forms of perception still is that, in picture perception, we are in a twofold state of seeing-in: We do not see just a concrete object, but a pictorial space encoded upon a surface. However, this discrimination between *face-to-face seeing* and

seeing-in (as well as between a surface and a pictorial space) is offered by ventral processing, which distinguishes between objects in the flesh (surfaces included), upon which motoric dorsal responses can be exercised, and depicted objects, with which no action is possible. This depends on the ventral (correct) encoding of the presence of a surface upon which the pictorial space is realized. Thus, ventral processing grants a twofold state of seeing-in.

This makes both objects very similar in the kingdom of perception, aligning to the adequacy of an experimental methodology using pictures to understand visual brain functions, also for visuomotor functions. Several stimuli employed in TVSM's experimental tests are indeed pictorial. Ventral vision represents the crucial differences between them.

The DVAPP also accounts for emotional responses to pictures (Ferretti 2017a). Each stream connects to a different portion of the orbito-frontal cortex,¹⁴ respectively, mediating the emotional evaluation of the stimulus visually encoded by each stream, concerning both recognition and visuomotor responses. This explains how we are emotionally involved with pictorial contents, but also how the surface properties can influence, emotionally, our perception of this content.

At this point, however, it remains to explain the nature of exotic perceptual cases *à la trompe l'oeil*, for since dorsal perception is involved in the perception of usual depicted objects, we cannot invoke any dorsal *exclusive* participation on pictorial space processing to explain what is different in this illusory case. This extends Nanay's (2015) DVAPP.

The answer, following the extension (Ferretti 2016a), may consider that ventral processing performs the recognition of what is a concrete object, like the surface, and what is a two-dimensional pictorial space. Thus, to explain *trompe l'oeils*, rather than “including *improper* dorsal processing” on the depicted object, we should “exclude *proper* ventral functioning” on the surface¹⁵. Why do we have the impression we can interact with the object (cf., OO above)? Because the surface cannot be tracked, and two-foldness is broken: We only see the depicted object, which does not look pictorial anymore. Why so? Because it is ventral processing that cannot detect the presence of any surface (masterfully hidden by the painter). Thus, a ventral representation responsible for tracking the presence of a surface is needed, to enter proper two-foldness for seeing-in, and to avoid a pictorial breakdown *à la trompe l'oeil*. Therefore, the representation of the surface we need must be ventral, not dorsal: it is about recognition in the first place. Dorsal processing (concerning OOO) cannot play any role here (cf. note ¹⁵) (though a dorsal representation of the action properties of the surface is still in play).

This would change the story concerning (b): we have a dorsal perception of the depicted object, that is, (b*). But also concerning the relation between (d) and (c): we need a representation of the surface, to avoid *trompe l'oeil* illusions,¹⁶ and this is ventral, not dorsal (cf. note ¹⁵). And concerning (c), about spatial relations and compensations: as dorsal processing responds also to depicted objects, and not only to concrete objects for effective interaction, like the surface, it may not be the one

discriminating about the different spatial relations between us and the two folds, performing compensation between the surface and the pictorial space (as indeed it does not fully discriminate between them). The response may relate to ventral processing, capable of such recognition.

All in all, the story would go, it is not that dorsal vision does not operate on usual pictorial spaces and then enters the game with *trompe l'oeils*. It is always there (both on the surface and on the pictorial space). Rather, it is ventral vision that should be always there, and leaves the room, with (non-trackable) surfaces, when there is a *trompe l'oeil*. Then, both streams can respond to pictorial arrangements, but only ventral processing can represent them as such, and distinguish between them and concrete objects (like surfaces), when not threatened by illusions (cf. note ¹⁵).

That said, to cash out the full answer, we need a further upgrade of the DVAPP, based on recent sets of evidence about the TVSM. This leads to understand the difference between objects in the flesh, pictorial objects, and *trompe l'oeil* illusions.

7 | A Further Extension of the Neural Dynamics of Seeing-In

The previous extension of the DVAPP (Ferretti 2016a) still operates endorsing functional difference between the streams. A further extension of it (Ferretti 2018b)¹⁷ offers a review of the most recent literature that goes beyond functional segregation between the streams, suggesting they work in interplay (cf. Section 3):

- i. Both streams can have an individual role in both recognition and vision-for-action.
- ii. The streams interact in many tasks, of recognition and vision-for-action, usually assigned, exclusively, to one stream.

Following (i), ventral vision can also be responsible for visually guided action, whereas dorsal vision can also be responsible for visual recognition. For example, ventral processing subserves *action planning*, concerning abstract aspects of an action: selecting the type of grip, the rotation of the wrist etc., (whereas dorsal processing subserves *motor programming*, concerning smooth parameters on grip force, hand closure, etc., computed in a specific situation) (Ferretti 2018b, 1303, 1304). Accordingly, dorsal processing subserves object recognition, especially for objects presented within peripersonal action space (Ferretti 2018b, 1291, 1292, 1296).

Following (ii), the streams interact in both object recognition and visually guided action. Thus, when evaluating visual tasks, instead of talking of purely ventral and dorsal processing, we should talk of an interplay between them. One in which ventral processing has a major role and dorsal processing a minor one (a mainly ventral interplay), and vice versa, one in which dorsal processing has a major role and ventral processing a minor one (a mainly dorsal interplay). For example, vision-for-action needs

both motor representations mainly driven by dorsal processing and semantic representations about objects' functions (a pen can be used to draw but also as a knife), offered by ventral processing (Ferretti 2018b, 1298).

Furthermore, the dorsal stream can be divided into a dorso-dorsal stream and a ventro-dorsal stream (cf. Section 3). The latter goes in the direction of the inferotemporal cortex related to the ventral stream and exhibits ventral computational features, managing many tasks related to visual recognition for action (Ferretti 2018b, 1291, 1292).

This evidence leads to further extend DVAPP's four claims, respecting Nanay's insights, while considering the addition about dorsal vision (Ferretti 2016a):

- A. The visual recognition of the depicted object, a function attributed to ventral vision (a), is shaped by (a mainly ventral) interstream interplay.
- B. The representation of action properties related to the depicted object, a function attributed to dorsal vision (b*), is shaped by (a mainly dorsal) interstream interplay.
- C. The representation of action properties related to the picture's surface, a function attributed to dorsal vision (c), is shaped by (a mainly dorsal) interstream interplay.
- D. The visual recognition of the picture's surface, a function attributed to ventral vision (d), is shaped by (a mainly ventral) interstream interplay.

Let us now appreciate the implications of this extension in detail. A crucial difference between Nanay's (2011, 2015) DVAPP and its first extension (Ferretti 2016a) is that only the latter suggests that pictorial space is represented by dorsal perception.

An inference from this first extension, as seen, shifts the explanation of *trompe l'oeils*, suggesting that ventral processing is responsible for tracking the surface, which with *trompe l'oeils* cannot be tracked, this leading to the illusion. The explanation does not invoke dorsal processing of the depicted object. This changes the relation between (b), (c), and (d). Concerning (b), dorsal processing represents the depicted object. Concerning (d), we need a representation of the surface to avoid *trompe l'oeil* illusions, and this is ventral, not dorsal, concerning (c) (cf. note ¹⁵).¹⁸

The new extension (Ferretti 2018b) suggests that *trompe l'oeil* illusory effects are explained considering a mainly ventral interplay, not just ventral processing. Ventral processing performs a recognition distinguishing between concrete objects, like the surface, and two-dimensional pictorial spaces. But this function is aided by dorsal processing's recognitional shade.

The first extension also impacts, as seen, other aspects of surface perception, *as per* (c). For Nanay's (2011, 2015) DVAPP, as the spatial relation of the observer with respect to the surface is different compared to the one with respect to the pictorial object, dorsal processing is responsible for the encoding of these spatial relations, and for the compensation thereof: dorsal processing,

managing action, has a special connection to space. Thus, it must represent the surface. This also grants two-foldness.

An inference from the first extension (Ferretti 2016a), as seen, is that ventral processing, not dorsal processing, represents the spatial aspects about the surface and their difference with the depicted object, with respect to the spatial relation of the observer, also concerning the compensation. This is the case for the reasons above exposed on its recognitional functions. Thus, we need a representation of the surface, subserving this spatial function, that is ventral, *as per* (d), not dorsal, *as per* (c).

The new extension suggests the mainly ventral interplay is responsible for encoding the surface's spatial qualities, and their difference with the pictorial space, with respect to the spatial relations with the observer, compensation included. Indeed, the mainly ventral interplay is the one that can track the presence of a concrete surface, across which are visually encoded two-dimensional marks visually represented as pictorial. This is strengthened by evidence that all vision is egocentric, and so also ventral vision is. Egocentric recognition is then crucial for these spatial aspects (Ferretti 2018b).

All in all, this new extension explains how the streams, working together, perform seeing-in, by granting two-foldness. And it also explains the breakdown of picture perception. Tracking the presence of a surface, and our spatial-motor relation to it, is crucial to perceive a pictorial space, granting two-foldness, and then seeing-in. When our visual system cannot track it, it is deceived that the unique object in front of it is the pictorial space, which, at that point, does not look pictorial anymore: if no surface is tracked, the visual system represents the pictorial object as an object in the flesh, that is, we fall into the *trompe l'oeil* effect. To offer this explanation, this further extension couples neural evidence on streams interplay with evidence from optics, psychophysics, and phenomenology that, when a surface is not visible, the perceiver enters the illusion of a three-dimensional object that offers interaction (Vishwanath 2014; Vishwanath and Hibbard 2013; Ferretti 2020a, 2020b). This multilevel analysis explains the relation between ordinary seeing-in, two-foldness, and *trompe l'oeil* effects (with respect to face-to-face visual perception).

If so, a mainly ventral interplay is crucial both for tracking the surface (unconsciously) and appreciating the depicted object (consciously), while avoiding illusory effects. *Ipsa facto*, it is crucial for distinguishing between pictures of objects and objects in the flesh. Indeed, it is by tracking a surface, which is an object in the flesh across which a pictorial space can be visually encoded, that the mainly ventral interplay can distinguish between an apple visually encoded on a surface (a surface that is a concrete object, offering reliable interaction) and a concrete apple in the flesh, offering reliable interaction (Ferretti 2018b, 4.4). Thus, the mainly ventral interplay manages all the spatial differences of the two folds with respect to the observer. The mainly dorsal interplay (unconsciously) represents action properties on both the surface and the depicted object, but this representation is handled by the results of a mainly ventral interplay. Crucially, with *trompe l'oeils*, we cannot track any surface, either consciously or unconsciously, and this disrupts two-foldness for seeing-in (Ferretti 2018b).

This extension preserves the aim by Nanay about coupling the notion of seeing-in, by Wollheim, with the TVSM. In this respect, as ventral processing, also the mainly ventral interplay can be conscious or unconscious, while the mainly dorsal interplay is most of the time unconscious, as dorsal processing. Thus, in correct seeing-in, we need our mainly ventral interplay consciously attuned onto the depicted object, while maintaining an unconscious mainly ventral interplay also on the surface, this granting its tracking as a concrete object for interaction, and thus avoiding *trompe l'oeil* troubles. At the same time, the mainly dorsal interplay, responsible for action responses, remains unconscious both concerning surface processing and the processing of the depicted object for the detection of action properties.

This also aligns to the notion of *inflection*, or inflected seeing-in, offered by Nanay, when considering esthetic appreciation (Nanay 2010a, 2011, 2016, 2017): it is the mainly ventral interplay that manages the recognition of the pictorial object and of the picture surface, as such, that is, as the vehicle of the pictorial object.¹⁹

8 | The Importance of the DVAPP

This further extension strengthens the marriage between Wollheim's theory and the TVSM, proposed by Nanay, through the DVAPP, with specific benefits for its theoretical power.

First, experimental results suggest that the visual mechanism subserving the recognition responsible for the impression of interactability with a concrete object is given by stereopsis, related to egocentric absolute depth spatial processing (see Ferretti 2016a, on Vishwanath 2014; see also Ferretti 2023, 2025). The second extension (Ferretti 2018b) suggested that, being stereopsis related to a recognition crucial for action, it comes from a mainly ventral interplay. Subsequent results suggested the collaboration of ventral and dorsal areas, via the posterior parietal cortex, in generating stereopsis (Uji et al. 2019). This confirms that this interplay is responsible for distinguishing between objects in the flesh and pictorial objects, as well as between pictorial spaces and surfaces, based on a recognition (Snow et al. 2011) that is useful to action (Freud et al. 2018) (cf. Uji et al. 2019), and that makes pictures and objects in the flesh both encoded (see Ferretti 2017b on Snow et al. 2011), though differently, from our visuomotor brain (see Ferretti 2018b, footnote 22, on Freud et al. 2015).

Second, this extension recognizes and explains a relation between *trompe l'oeils* and *anamorphosis*. With standard depicted objects, we do not perceive interactability (as with *trompe l'oeils*) and sensorimotor perspectival changes (as with *anamorphosis*) (Matthen 2005; Nanay 2010b, 2011, 2015, 2017; Vishwanath et al. 2005). However, *trompe l'oeils* also seem to display *anamorphosis*, then fostering both perceptual effects (Ferretti 2020a, 2020b). The extension suggests that, with non-illusory pictures, a mainly ventral interplay is responsible for the correct encoding of a surface, and its spatial properties, related to sensorimotor perspectival changes and interactability. This prevents these perceptual effects from occurring with respect to the pictorial

space. Conversely, with *trompe l'oeils*, this interplay loses reference to a trackable surface. Then, these perceptual effects are allocated to the pictorial space, which does not look pictorial anymore, offering sensorimotor perspectival changes and interactability. Therefore, the spatial relations within the two-folds, with respect to the observer's spatial, bodily and motor perspective, are granted by a mainly ventral interplay. We can also escape *trompe l'oeil* illusions while moving with respect to the center of projection of the picture (the perspective from which the illusion works out). This perspectival change in recognition is due to a mainly ventral interplay. Thus, a break in this interplay disrupts the equilibrium of the spatial relations between the two folds with respect to the viewer (Ferretti 2021).

If so, this involvement of an action dimension on recognition, emerging from this interplay, suggests *why the pictorial needs the motoric* (Ferretti 2021), demanding an *action-based theory of picture perception*. When action processing is correctly attuned to the surface, we have pictorial experience. When not, pictoriality is lost: We fall victim of *trompe l'oeil* effects with *anamorphosis*. Therefore, we need a “motoric two-foldness”, to obtain a “motoric seeing-in”. Otherwise, we may become blind to seeing-in, or “blind-in”, losing the experience of the picture.

Furthermore, this extension offers a grammar of spaces unifying the DVAPP along its different versions. It distinguishes (Ferretti 2018b, 3.1) between peripersonal space, that is, the space of action, egocentric space, that is, the observer's point of view, and sensorimotor perspectival changes, that is, changes in visual experience brought by movement, this considering all vision as egocentric, even admitting viewer-variant and viewer-invariant visual processing (suggesting all peripersonal space is egocentric, but not vice versa). This makes egocentricity (Matthen 2005, 2021; Nanay 2011, 2015), peripersonal space (Ferretti 2016a), sensorimotor changes (Nanay 2010b, 2011, 2015), and egocentric absolute depth in stereopsis (Vishwanath 2014; Ferretti 2018b, 2023) all compatible.

This story, started with Matthen and Nanay, following the insights by Wollheim, makes justice of the results from visual neuroscience, accounting for *the neural dynamics of seeing-in* (cf. the title of Ferretti 2018b), by coupling the recent results from the TVSM with DVAPP's insights. The distinction between a mainly ventral and a mainly dorsal interplay in object recognition and visual guidance of action extends Nanay's story concerning the ventral/dorsal distinction, while preserving its main Wollheimian insights, and is compatible with Matthen's distinction between motion-guiding vision and descriptive vision. This story offers a unified account of the neural dimension, the psychological dimension, and the phenomenological dimension, of picture perception, with respect to face-to-face perception, and to the puzzling case of *trompe l'oeil* illusions.

Acknowledgments

I thank two anonymous Reviewers. This work was supported by a German Humboldt Fellowship, hosted by Professor Albert Newen at the Institute for Philosophy II, Ruhr-University Bochum, Germany. Open

access publishing facilitated by Università degli Studi di Bergamo, as part of the Wiley - CRUI-CARE agreement.

Funding

The author has nothing to report.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

The author has nothing to report.

Endnotes

- ¹ There is a debate on whether this awareness in pictures implies attention, with respect to consciousness (Ferretti and Marchi 2020, 2024).
- ² See (Hopkins 2012; Nanay 2017; Ferretti and Marchi 2020).
- ³ This excludes visual alternation between the surface and the pictorial space (Gombrich 1960; Ferretti 2018a).
- ⁴ *Trompe l'oeils* lead to a breakdown of seeing-in, fostering visual experience of an object in the flesh. *Hyper-realist* pictures do still lead to seeing-in, but with more enhanced details with respect to usual pictures (Ferretti 2018b, footnote 30).
- ⁵ This is compatible with Matthen's use of a functional label (motion-guiding-vision/descriptive vision) rather than a neurophysiological one (dorsal/ventral).
- ⁶ There is a debate on inflection, as simultaneous consciousness of both the surface and the pictorial space may be problematic (Hopkins 2012; Nanay 2017; Ferretti and Marchi 2020). However, it seems that ventral conscious recognition represents the pictorial space, whereas ventral *unconscious* recognition represents the surface.
- ⁷ “When we see *trompe l'oeil* paintings and are deceived by them, our ventral stream still represents the depicted objects, as in the case of other instances of picture perception. But, and here is the difference between seeing *trompe l'oeil* paintings and perceiving other pictures, our dorsal stream also represents the depicted objects. We have seen that normally, the dorsal stream does not represent the depicted object” (Nanay 2015, 193).
- ⁸ Even if not all of them, as there is, for example, no massive depth visual experience of those depicted objects we recognize; except, of course, for *trompe l'oeil* cases.
- ⁹ Referred, sometimes, to affordances in pictures (Zipoli Caiani 2013).
- ¹⁰ Depicted objects can be apparently located within the peripersonal action space of the observer.
- ¹¹ What permits discrimination on these different objects flags, accordingly, what permits discrimination between the theories.
- ¹² Dorsal vision can encode action properties recalled by the shape represented by the depicted object.
- ¹³ The spatial property of a mug (it is wavy handle) can be represented, dorsally, as an action property (the property of being grasped), *as per* (OOO). This may open to an effective action possibility (graspability), whose evaluation, based on recognition, is ventrally performed, *as per* (OO). This is a crucial functional distinction, concerning vision and action in relation to the streams, within this extension of the DVAPP (Ferretti 2016a, 4.1; 2018b, 3.2).
- ¹⁴ The dorsal stream is connected to the medial orbito-frontal cortex, whereas the ventral stream is connected to lateral orbito-frontal cortex.

¹⁵ One move for Nanay's DVAPP to explain *trompe l'oeils* would be invoking not only dorsal processing on the depicted object (cf. note ⁷) but also lack of dorsal processing on the surface. This is not considered in that formulation: dorsal processing would still be attuned also to the surface (Nanay 2015, 194). (Arguably, because representing the surface has no such an important role as in later formulations of the DVAPP). Nanay also suggests that, with *trompe l'oeils*, we presumably do not ventrally see the surface. But this is, arguably, because there is no inflection: Lack of ventral processing on the surface is not taken as a piece of explanation of the illusion (as it is for DVAPP's extension), only dorsal attunement on the depicted object is (2015, 194). Ventral processing of the surface does not play a crucial role in Nanay's DVAPP, as it could or could not be at work, that is, (d). The main role on the surface is played by dorsal processing, that is, (c).

¹⁶ More precisely, we have a ventral conscious representation of the depicted object and a ventral unconscious representation of the surface.

¹⁷ But see some germinal idea on this position in (Ferretti 2016b).

¹⁸ Cf. note ¹⁶.

¹⁹ For the problem of simultaneous consciousness and inflection, see (Ferretti and Marchi 2020).

References

Farah, M. 2004. *Visual Agnosia*. 2nd ed. MIT Press.

Ferretti, G. 2016a. "Pictures, Action Properties and Motor Related Effects." *Synthese, Special Issue: Neuroscience and Its Philosophy* 193, no. 12: 3787–3817. <https://doi.org/10.1007/s11229-016-1097-x>.

Ferretti, G. 2016b. "Visual Feeling of Presence." Supplement, *Pacific Philosophical Quarterly* 99, no. S1: 112–136. <https://doi.org/10.1111/papq.12170>.

Ferretti, G. 2017a. "Pictures, Emotions, and the Dorsal/Ventral Account of Picture Perception." *Review of Philosophy and Psychology* 8, no. 3: 595–616. <https://doi.org/10.1007/s13164-017-0330-y>.

Ferretti, G. 2017b. "Are Pictures Peculiar Objects of Perception?" *Journal of the American Philosophical Association* 3, no. 3: 372–393. <https://doi.org/10.1017/apa.2017.28>.

Ferretti, G. 2018a. "The Nature of Pictorial Representations." *Phenomenology and Mind* 14: 136–144. https://doi.org/10.13128/Phe_Mi-23631.

Ferretti, G. 2018b. "The Neural Dynamics of Seeing-In." *Erkenntnis* 84, no. 6: 1285–1324. <https://doi.org/10.1007/s10670-018-0060-2>.

Ferretti, G. 2020a. "Why Trompe l'oeils Deceive Our Visual Experience." *Journal of Aesthetics and Art Criticism* 78, no. 1: 33–42. <https://doi.org/10.1111/jaac.12688>.

Ferretti, G. 2020b. "Do Trompe l'oeils Look Right When Viewed From the Wrong Place?" *Journal of Aesthetics and Art Criticism* 78, no. 3: 319–330. <https://doi.org/10.1111/jaac.12750>.

Ferretti, G. 2021. "Why the Pictorial Needs the Motoric." *Erkenntnis* 88, no. 2: 771–805. <https://doi.org/10.1007/s10670-021-00381-1>.

Ferretti, G. 2023. "For an Epistemology of Stereopsis." *Review of Philosophy and Psychology* 16, no. 1: 267–284. <https://doi.org/10.1007/s13164-023-00711-y>.

Ferretti, G. 2025. "What Stereoblindness Teaches Us About Visual Reality." *Phenomenology and the Cognitive Sciences*: 1–29. <https://doi.org/10.1007/s11097-025-10054-x>.

Ferretti, G., and F. Marchi. 2020. "Visual Attention in Pictorial Perception." *Synthese* 199, no. 1–2: 2077–2101. <https://doi.org/10.1007/s11229-020-02873-z>.

Ferretti, G., and F. Marchi. 2024. "Trompe l'œil Illusions: Pay (Visual) Attention." *Argumenta* 1–22. <https://doi.org/10.14275/2465-2334/20240.fer>.

Ferretti, G., and S. Zipoli Caiani. 2024. "An All-Purpose Framework for Affordances. Reconciling the Behavioral and the Neuroscientific Stories." *Synthese, Special Issue: Neuroscience and Its Philosophy* 204, no. 1: 25. <https://doi.org/10.1007/s11229-024-04646-4>.

Freud, E., T. Ganel, I. Shelef, M. D. Hammer, G. Avidan, and M. Behrmann. 2015. "Three-Dimensional Representations of Objects in Dorsal Cortex Are Dissociable From Those in Ventral Cortex." *Cerebral Cortex* 27: 1–13. <https://doi.org/10.1093/cercor/bhv229>.

Freud, E., S. N. Macdonald, J. Chen, D. J. Quinlan, M. A. Goodale, and J. C. Culham. 2018. "Getting a Grip on Reality: Grasping Movements Directed to Real Objects and Images Rely on Dissociable Neural Representations." *Cortex* 98: 34–48. <https://doi.org/10.1016/j.cortex.2017.02.020>.

Gombrich, E. 1960. *Art and Illusion*. Pantheon.

Hopkins, R. 1998. *Picture, Image and Experience. A Philosophical Inquiry*. Cambridge University Press.

Hopkins, R. 2012. "Seeing-In and Seeming to See." *Analysis* 72, no. 4: 650–659. <https://doi.org/10.1093/analysis/ans119>.

Jacob, P., and M. Jeannerod. 2003. *Ways of Seeing. the Scope and Limits of Visual Cognition*. Oxford University Press.

Jeannerod, M. 2006. *Motor Cognition: What Actions Tell the Self*. Oxford University Press.

Matthen, M. 2005. *Seeing, Doing and Knowing: A Philosophical Theory of Sense Perception*. Oxford University Press.

Matthen, M. 2010. "Two Visual Systems and the Feeling of Presence." In *Perception, Action, and Consciousness: Sensorimotor Dynamics and the Two Visual Systems*, edited by N. Gangopadhyay, M. Madary, and F. Spencer, 107–124. Oxford University Press.

Matthen, M. 2021. "Objects, Seeing, and Object-Seeing." *Synthese* 198, no. 4: 3265–3288. <https://doi.org/10.1007/s11229-019-02279-6>.

Milner, A. D., and M. A. Goodale. 1995/2006. *The Visual Brain in Action*. 2nd ed. Oxford University Press.

Nanay, B. 2010a. "Inflected and Uninflected Experience of Pictures." In *Philosophical Perspectives on Depiction*, edited by C. Abell and K. Bantinaki, 181–207. Oxford U.P.

Nanay, B. 2010b. "Transparency and Sensorimotor Contingencies: Do We See Through Photographs?" *Pacific Philosophical Quarterly* 91, no. 4: 463–480. <https://doi.org/10.1111/j.1468-0114.2010.01380.x>.

Nanay, B. 2011. "Perceiving Pictures." *Phenomenology and the Cognitive Sciences* 10, no. 4: 461–480. <https://doi.org/10.1007/s11097-011-9219-x>.

Nanay, B. 2015. "Trompe l'oeil and the Dorsal/Ventral Account of Picture Perception." *Review of Philosophy and Psychology* 6, no. 1: 181–197. <https://doi.org/10.1007/s13164-014-0219-y>.

Nanay, B. 2016. *Aesthetics as Philosophy of Perception*. Oxford University Press.

Nanay, B. 2017. "Threefoldness." *Philosophical Studies* 175, no. 1: 163–182. <https://doi.org/10.1007/s11098-017-0860-2>.

Rossetti, Y., L. Pisella, and R. D. McIntosh. 2017. "Rise and Fall of the Two Visual Systems Theory." *Annals of Physical and Rehabilitation Medicine* 60, no. 3: 130–140. <https://doi.org/10.1016/j.rehab.2017.02.002>.

Sakreida, K., I. Effner, S. Thill, et al. 2016. "Affordance Processing in Segregated Parieto-Frontal Dorsal Stream Sub-Pathways." *Neuroscience & Biobehavioral Reviews* 69: 89–112. <https://doi.org/10.1016/j.neubiorev.2016.07.032>.

Snow, J. C., C. E. Pettypiece, T. D. MacAdam, et al. 2011. "Bringing the Real World Into the fMRI Scanner: Repetition Effects for Pictures Versus Real Objects." *Scientific Reports* 1, no. 1: 130. <https://doi.org/10.1038/srep00130>.

- Uji, M., A. Lingnau, I. Cavin, and D. Vishwanath. 2019. "Identifying Cortical Substrates Underlying the Phenomenology of Stereopsis and Realness: A Pilot fMRI Study." *Frontiers in Neuroscience* 13: 646. <https://doi.org/10.3389/fnins.2019.00646>.
- Ungerleider, L., and M. Mishkin. 1982. "Two Cortical Visual Systems." In *Analysis of Visual Behavior*, edited by D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, 549–586. MIT Press.
- Vishwanath, D. 2014. "Toward a New Theory of Stereopsis." *Psychological Review* 121, no. 2: 151–178. <https://doi.org/10.1037/a0035233>.
- Vishwanath, D., A. R. Girshick, and M. S. Banks. 2005. "Why Pictures Look Right When Viewed From the Wrong Place." *Nature Neuroscience* 8, no. 10: 1401–1410. <https://doi.org/10.1038/nn1553>.
- Vishwanath, D., and P. Hibbard. 2013. "Seeing in 3D With Just One Eye: Stereopsis in the Absence of Binocular Disparities." *Psychological Science* 24, no. 9: 1673–1685. <https://doi.org/10.1177/0956797613477867>.
- Wollheim, R. 1980. "Seeing-As, Seeing-In, and Pictorial Representation." In *Art and Its Object*. 2nd ed., 205–226. Cambridge University Press.
- Wollheim, R. 1998. "On Pictorial Representation." *Journal of Aesthetics and Art Criticism* 56, no. 3: 217–226. https://doi.org/10.1111/1540_6245.jaac56.3.0217.
- Zipoli Caiani, S. 2013. "Extending the Notion of Affordance." *Phenomenology and the Cognitive Sciences* 13, no. 2: 275–293. <https://doi.org/10.1007/s11097-013-9295-1>.