

Triangulating Depiction: Pictorial Experience, Vision Science, and the Standard of Correctness

Luca Marchetti

University of Genoa

Abstract

Smith's *Film, Art and the Third Culture* is dedicated to developing a naturalized aesthetics of film, and at the heart of his proposal for a methodological "triangulation" is the principle that our research should take serious account of three levels of analysis, each with its respective type of evidence: the phenomenological level, the psychological level, and the neurophysiological level. While Smith addresses many aspects of our perceptual, emotional, and cognitive engagement with movies within this framework, he rarely discusses what underlies our visual engagement with films: the fact that they are depictions, i.e. (moving) pictorial representations. In this paper, I aim to fill this gap by discussing the notions of depiction and pictorial experience within Smith's methodological framework. In the first part, I analyse the relationship between empirical research in experimental psychology and neurocognitive sciences and the philosophy of pictures through the lens of triangulation, showing there are good empirical reasons to hold Wollheim's seeing-in theory of pictorial experience. In the second part, I emphasise that pictures are cultural products grounded in authorial intentions, conventions, and norms, and I propose a way in which a naturalized account could shed light on the relationship between a subset of these norms—what Wollheim termed "the standard of correctness"—and pictorial experience.

Keywords: Depiction, Pictorial experience, Seeing-in, Standard of correctness, Perception, Cognitive penetration, Cognitive phenomenology.

1. Introduction

Smith's *Film, Art and the Third Culture* is dedicated to developing a naturalized aesthetics of film, and at the heart of his proposal for a methodological "triangulation" is the principle that our research should take serious account of three levels of analysis, each with its respective type of evidence: the phenomenological level, the psychological level, and the neurophysiological level. While Smith addresses many aspects of our perceptual, emotional, and cognitive engagement with movies within this framework, he rarely discusses what underlies our visual engagement with films:

the fact that they are depictions, i.e. (moving) pictorial representations. In this paper, I aim to fill this gap by discussing the notions of depiction and pictorial experience within Smith’s methodological framework. In Section 2, I analyse the relationship between empirical research in experimental psychology and neurocognitive sciences and the philosophy of pictures through the lens of triangulation, showing there are good empirical reasons to hold Wollheim’s seeing-in theory of pictorial experience. In Section 3, I emphasise that pictures are cultural products grounded in authorial intentions, conventions, and norms, and I propose a way in which a naturalized account could shed light on the relationship between a subset of these norms—what Wollheim termed “the standard of correctness”—and pictorial experience.

2. Pictorial Experience, Triangulated

The debate on depiction—how we should characterise the peculiar mode of representation that distinguishes pictures from other kinds of representations—and the debate on the distinctive kind of experience pictures elicit—what it is like to experience pictures—are largely philosophical. Yet, since the publication of Gombrich’s *Art and Illusion* in 1961, and continuing into contemporary discussions, some philosophers have drawn on empirical data from the sciences of the mind to support specific claims about depiction, depicted properties, and pictorial experience.¹ While this approach is promising and has enhanced our understanding of both depiction and pictorial experience, it is far from being the consensus view. The aim of this first section is to analyse the relationship between empirical research in experimental psychology and the neurocognitive sciences and the philosophy of pictures through the lens of the notion of triangulation. In section 1.1, I discuss the phenomenology of picture perception, with particular focus on Wollheim’s seeing-in theory (which Smith appears to endorse in his book). Then, in section 1.2, I examine psychological and neuroscientific data that appear to support the seeing-in theory. The upshot will be to stress how the method of triangulation proves fruitful in providing a plausible and comprehensive account of pictorial experience in terms of seeing-in.

2.1. Phenomenology

Experiential theories of depiction define pictorial representations as the kinds of things that generate peculiar kinds of experiences: pictures are created objects that have the function of generating a perceptual experience.² The experience elicited by pictures—usually visual (but for non-visual pictorial experience see e.g. Lopes 1997)—is a composite perceptual experience, in which the “perception” of the depicted scene (which is not in front of us) is generated by and experienced along the perception of the marked surface (the object that is actually in front of us). This experience is usually called a “pictorial experience”. In particular, for these theories, images have the function of generating a composite experience, one in which at the perception of a bidimensional marked surface—which is there—it is added the perception of a three-dimensional scene—which is not there. How should we characterize such an experience?

¹ See e.g. Briscoe 2016, Ferretti 2018, 2021a, 2021b, Marchetti 2022, Nanay 2011, 2018, Newall 2015, Zeimbekis 2015.

² In this sense, they are “experiential artifacts”. See Marchetti 2024 and Terrone 2024.

There are two main competing theory for the phenomenology of pictorial experience, Gombrich's and Wollheim's.³ For Gombrich (1960), images generate an alternation of simple perceptions, the veridical perception of the marked surface and the illusory perception of the depicted scene. In Gombrich's view, pictorial experience is the experience of seeing the picture as its subject. Indeed, he compares picture perception with the phenomenon of aspect-switching or "seeing as". As in the perceptually ambiguous duck-rabbit picture one can see that figure either as a duck or as a rabbit but never experience the two aspects at the same time, in pictorial experience either one sees the picture as a marked surface, or one sees the picture as a depicted scene. For Gombrich, the experience of the marked surface and the experience of the depicted scene are incompatible. Richard Wollheim (1980) attacks Gombrich's view of pictorial experience on exactly this ground: it is not the case that a spectator can only alternate between the experience of the surface and the experience of the content it represents. On the contrary, what is distinctive about pictorial experience is exactly the simultaneous awareness of the marks on a picture's surface and the scene it represents. In these terms, pictorial experience is defined by Wollheim as a "seeing-in" experience—when we face a picture, we see the picture's subject in a marked surface. And seeing-in experiences—that are not confined to pictures: e.g. we can see landscapes or faces in the stains on an old wall—are experiences with a *sui generis* phenomenology, which he calls "twofoldness". A viewer looking at a picture, Wollheim maintains, undergoes a "twofold" experience: on one hand, she is visually aware of the flat surface of the picture; on the other, she experiences the subject matter of the picture. Wollheim called the first of these folds of seeing-in the 'configurational' fold, and the second the 'recognitional' fold, and claims that these are two distinguishable but "inseparable" aspects of a single visual experience, and not two experiences somehow combined (Wollheim 2003).⁴

Smith, in one of the rare passages in which he discusses depiction and pictorial experience, seems to endorse Wollheim's theory. He writes, discussing on the notion of expansionism—i.e. the idea that, in our engagement with works of art,

³ Both theories have been developed and refined, but I use them here in a general sense to illustrate the two main camps. For the development of Gombrich's theory in relation to vision science, see Briscoe 2018; for Wollheim, see the next fn. 5.

⁴ As Voltolini (2015: 74) remarks, "according to Wollheim, over and above ascertaining the folds' inseparability, nothing more can be said in order to qualify the seeing-in experience, which therefore remains utterly ordinary on the one hand and quite mysterious on the other". While a detailed analysis of these issues lies outside the scope of this paper, let me briefly note that, following Wollheim, two main strands of theory have attempted to define this experience more precisely—partly because, as I will suggest, triangulation may help discard one of them (see fn. 5). These are, on the one hand, experienced resemblance theories (e.g. Peacocke 1987; Hopkins 1998), and on the other, make-believe theories (Walton 1990). Robert Hopkins (1998), who arguably develops the most refined version of an experienced resemblance account, maintains that seeing-in is an experience of resemblance in *outline shape* between the marks on the surface of a picture and the objects or scene they represent. An alternative way of building on Wollheim's seeing-in relies on imagination and is proposed by Kendall Walton (1990). In Walton's view, pictorial experience consists in imagining our actual seeing the marked surface—the representational prop—to be a face-to-face seeing of the depicted scene. For the empirical problems raised by Walton's theory see fn. 6 below.

our capacities are frequently extended in entirely new directions, operating in ways not seen outside artistic contexts⁵—that:

Arguably, seeing-in itself is the basic form of expansion here. Wollheim defines ‘seeing-in’ as our capacity to see, in a non-illusionary fashion, three-dimensional spaces and objects in two-dimensional surfaces, contrasting such ‘seeing-in’ with ordinary ‘face-to-face’ seeing. Although seeing-in is a natural perceptual capacity, evident in our ability to see faces and figures in clouds and other naturally occurring visual phenomena, the practices of still and moving depiction create a domain of experience quite distinct from face-to-face vision. In short, picture perception, and motion picture perception, are phenomenologically distinct from ordinary perception (Smith 2017: 42).

Smith does not elaborate further. In the second part of this first section, I argue that he is right to endorse seeing-in, and I support this claim by drawing on his own methodology—triangulation. Following Smith’s suggestion that, to have a proper understanding of aesthetic experiences—and of mental phenomena in general—we should use a triangulation between three bodies of data—phenomenological, psychological, and neurophysiological—the next subsection examines psychological and neuroscientific data that empirically ground the seeing-in theory.⁶ Seeing-in, these data show, is a natural perceptual capacity—it depends on our innate ability to perceive depth from monocular cues and to segregate visual information. However, it remains distinct from face-to-face vision in that surface and content are separately processed and integrated.

2.2. Psychology and Neurophysiology

Seeing-in is a natural perceptual capacity, it depends on our innate ability to perceive depth from monocular cues and to segregate perceptual cues. Robert Briscoe

⁵ In particular, Smith writes: “the character of many of the perceptual experiences generated by films and other artworks will often be quite unlike ordinary perceptual experience—not because of some ghostly or ‘non-natural’ sphere into which such experience propels us, but because artworks by design push and pull our perception in ways that happen rarely and only accidentally (if at all) in ordinary experience. The precisely engineered, undulating wave patterns of Bridget Riley’s Op Art paintings do not turn up in nature, any more than do the mosaic surfaces of Impressionism or the precisely rendered forms of realism. Even when motivated by realism, pictorial invention invariably delivers new forms of visual experience, rather than merely staging illusions of familiar experience” (Smith 2017: 37–38). For further discussion on how the notion of expansionism can be developed in relation to pictures, see fn. 17 below.

⁶ My strategy here is thus to examine how psychological and neuroscientific evidence can be seen to support an experiential theory, the idea being that any viable theory must align with the empirical data it seeks to explain or interpret (Voltolini 2013: 45). Another way to use psychological data, however, is not to support a theory, but to discard one. On this ground, data from developmental psychology can be used in order to criticize Walton’s idea that seeing-in images involve a special kind of make-believe (see fn. 4 above). Since Walton’s theory accounts for depiction by appealing to experiences that involve make-believe, it should follow that individuals that are unable to make believe cannot have such experiences. But developmental psychology shows that children may undergo “seeing-in” experiences at least three months before their manifestation of make-believe understanding (at the age of 8 months rather than at the age of 15). From an empirical point of view it is therefore doubtful that the experience of the depicted scene has to be interpreted as a make-believe perception, as Walton wishes (Voltolini 2013).

(2016) argues that perceptual psychology and psychophysics support what he calls the “Continuity Hypothesis”—the view that face-to-face perception and pictorial perception belong to the same psychological kind. Although this hypothesis has roots in Gombrich’s writings on depiction, it is far from a consensus view. Drawing on empirical studies of pictorial experience and the methodology of virtual psychophysics in vision science, Briscoe maintains that “just as much can be said about what the light reflected from a 2D surface must be like to elicit an experience as of a certain 3D spatial layout as can be said about what the light reflected from the world must be like in order to have this effect. To the extent that pictures and real-world scenes can cause the same types of information-bearing, proximal stimulation in their viewers, they can also cause their viewers to form perceptual states of the same psychological kind” (Briscoe 2016: 61–62). There is, in fact, substantial evidence that pictorial experience and face-to-face perception are supported by the same bottom-up sources of optical information and governed by the same computational principles. In particular, so-called pictorial cues—depth cues that are available even to one eye alone—play a central role in both kinds of experience. These include familiar phenomena such as occlusion (when one object partly covers another, suggesting that it is in front), linear perspective (the convergence of parallel lines toward a vanishing point), texture gradients (the progressive decrease in the size and density of surface textures with increasing distance), shading and lighting (which can indicate the shape and orientation of objects), and aerial perspective (the haziness or bluish tint of distant objects due to atmospheric scattering). These cues, when present in images, can trigger perceptual mechanisms that the visual system normally employs when processing three-dimensional spatial layout in the real world. The fact that these same mechanisms are activated in response to two-dimensional images supports the idea that pictorial perception and ordinary seeing share the same basic psychological architecture. It is precisely because these cues can simulate the patterns of light reflected from three-dimensional scenes that they can elicit robust experiences of depth and volume in the absence of actual spatial structure (see also Zeimbekis 2015).⁷ “[P]erceiving depth in pictures and perceiving depth in the real world,” as Cutting summarizing the empirical evidence concludes, “are cut from the same informational cloth” (2003: 236; cited in Briscoe 2016: 57). This assessment, Briscoe emphasizes, also extends to surface properties that are visually estimated, in part, on the *basis* of information specifying spatial layout (Fleming and Anderson 2004). The apparent lightness (albedo) of an object in pictorial space, for example, can be shifted from black to white by varying the perceived organization of the 3D scene in which it is located (Gilchrist 1977, 1980; Anderson and Winawer 2005). Briscoe notes that psychophysical evidence that the spatial contents of pictorial experience and seeing face-to-face vary, in a closely corresponding way, with changes in the bottom-up, optical information available to the eyes indicates that the visual system is processing that information—in both cases—in accordance with the

⁷ Movies make also available potent movement-based sources of distance and shape information, such as motion parallax and the kinetic depth effect (Wallach and O’Connell 1953; Ullman 1979). Some other kinds of pictures—for example, stereograms (on which VR headsets are based)—also take advantage of binocular sources of optical information, such as retinal disparity (Marchetti 2025b). Mirrors make available all the depth cues of the real world (Marchetti 2025a). While it can be argued that mirrors, VR, and even 3D movies do not elicit seeing-in experiences precisely because of this (Marchetti 2025a, b), traditional movies do (see Hopkins 2008)—as it is also assumed by Smith himself.

same set of computational formation principles. It thus provides strong support for the view they are experiences of the same psychological kind (i.e. the Continuity Hypothesis). As Briscoe argues, in addition to sources of psychophysical evidence, there is also neuropsychological support for the Continuity Hypothesis: pictures of 3D objects activate the same processing areas in the visual brain as are activated by the depicted objects themselves.⁸

There are continuities between face-to-face vision and seeing pictures, but also differences. From a neuropsychological point of view, seeing pictures remains distinct from face-to-face vision in that surface and content are processed separately and then integrated. This, I now show, is precisely what grounds, neuropsychologically, the phenomenology of seeing-in—its *sui generis* twofold character. Indeed, while Wollheim’s theory says little about how surfaces support the *sui generis* twofold experience, some psychologists have recently suggested how this might happen (see also Kulvicki 2013). As psychologists Reinhard Niederée and Dieter Heyer (2003) have shown, pictures present conflicting cues to our visual system, which must be resolved in order to sustain the experience of *seeing-in*. Specifically, monocular cues such as shading, perspective, and occlusion suggest the presence of depth and three-dimensional structure within the image, while binocular disparity and focus cues from stereopsis consistently indicate a flat, two-dimensional surface.⁹ Under these circumstances, Niederée and Heyer argue that the visual system does not merely reconcile the inconsistency, but actively segregates the cues into what they term subpercepts—perceptual representations of distinct objects or layers: on the one hand, the physical, marked surface of the picture; on the other, the represented content or the “virtual” scene. These subpercepts are not generated independently or in a sequential manner. Rather, they interact dynamically across various levels of perceptual processing, both bottom-up and top-down. In their model, the visual system supports a parallel representation of surface and content by allowing feature-based cross-talk between the two levels. This layered yet integrated structure provides a plausible psychological basis for the phenomenological duality central to Wollheim’s notion of *seeing-in*—the simultaneous awareness of a marked surface and a depicted scene.

Seeing-in (and the cue segregation hypothesis) can also be supported by neurophysiological evidence—the third angle of Smith’s triangulation. Several philosophers, including Mohan Matthen (2005, chap. 13), Bence Nanay (2011), and

⁸ This is true, Briscoe argues, not only of the object-categorizing ventral stream (see Grill-Spector et al. 2000 and Kanwisher 2004, e.g., for brain-imaging studies on how ventral stream areas respond to pictures of different kinds of objects), but also of the action-guiding, dorsal stream. Neurons in the intraparietal sulcus, a dorsal stream area involved in the skillful guidance of eye movements as well as reaching and grasping, for example, selectively respond to objects in pictorial space with different texture gradient and linear-perspective-defined 3D shapes and orientations (Taira 2001; James et al. 2002; Tsutsui et al. 2002, 2005; Sakata et al. 2003; Nelissen et al. 2009). There is also evidence that neurons in area V3a are sensitive to depth in photographs (Berryhill and Olson 2009) and that neurons located in the lateral occipito-parietal junction, another early dorsal stream area, distinguish between images of graspable and nongraspable objects (Rice et al. 2007). Consistent with these findings, behavioral studies indicate the grasping system in the dorsal stream is able to exploit pictorial depth cues, although binocular cues are more heavily weighted, especially when grasping with the right hand (Marotta and Goodale 2001; Westwood et al. 2002; Gonzalez et al. 2008). On this, see also Ferretti 2021a, b.

⁹ See also fn. 6.

Gabriele Ferretti (2018), building on the framework of the dorso-ventral hypothesis, have proposed that objects in pictorial space are represented by the object-categorising ventral processing stream, but not typically by the action-guiding dorsal stream.¹⁰ This offers another way of providing empirical support for the twofoldness thesis, drawing from neuroscience rather than perceptual psychology. I take Nanay's account as an illustrative example. Nanay's central claim is that the characteristic "twofoldness" of picture perception—originally described by Wollheim as the simultaneous awareness of both the depicted scene and the picture surface—can be given a neurofunctional explanation. According to him, when we perceive a picture, our ventral stream attributes properties to the depicted scene, while our dorsal stream attributes properties to the surface of the picture. In particular, Nanay argues for the following four claims: (a) the depicted scene is represented by ventral perception, (b) the depicted scene is not represented by dorsal perception, (c) the picture surface is represented by dorsal perception, and (d) the picture surface is not necessarily represented by ventral perception. These claims are substantiated through a wide range of neuroscientific findings and experimental data. First, to support his claim that the ventral stream is responsible for recognising the depicted scene in a picture, Nanay turns to neuropsychological evidence from cases of visual agnosia—disorders involving damage to the ventral stream. He references studies by Turnbull et al. (2004) and Westwood et al. (2002), which show that patients with visual agnosia are unable to recognise objects in pictures despite being able to copy lines and shapes accurately.¹¹ Second, Nanay argues that the dorsal stream does not represent the depicted scene, because it is specialised for egocentric localisation and the visual guidance of action, both of which require access to the actual spatial layout of the environment. Since depicted objects are not in the perceiver's egocentric space, they cannot be localised by the dorsal stream.¹² Third, in arguing that the picture surface is represented by dorsal perception, Nanay revisits a range of empirical findings related to perceptual constancy when viewing pictures from oblique angles. It is well

¹⁰ The ventral stream, sometimes referred to as the "what" pathway, is responsible for object identification and recognition, while the dorsal stream, or "how/where" pathway, is involved in guiding motor actions based on spatial information. This distinction, based on both anatomical and functional data, has been thoroughly documented by Milner and Goodale (1995), Goodale and Milner (2004), and Jeannerod (1997). Importantly, Nanay remains neutral on contentious issues such as whether the dorsal stream operates unconsciously (Milner and Goodale 1995; critiqued by Dehaene et al. 1998 and Jeannerod 1997) or whether it processes nonconceptual content (Clark 2001), focusing instead on the basic functional separation between the two systems.

¹¹ For instance, the patient D.M. could replicate the structure of a drawing but could not identify what it depicted or whether it showed an impossible object. Such patients are also not subject to standard pictorial illusions like the Müller-Lyer and Ponzo illusions, which normally depend on interpreting two-dimensional images as three-dimensional scenes. These deficits point to the necessity of ventral processing for interpreting depicted content.

¹² He draws here on Matthen's (2005: 312–315) argument that pictorial space gives us no information about the actual location of depicted objects. Further support comes from classical philosophical accounts of egocentric localisation by Evans (1982, 1985) and Poincaré (1905/1958), who conceptualise egocentric space in terms of possible movements or action plans. Nanay supplements these accounts with the sensorimotor view advocated by Noë (2002, 2003, 2004) and Campbell (2011), arguing that localisation involves not only representing distance but also expecting how this distance would change with movement. Since we cannot interact with depicted objects as if they were physically present, he concludes that dorsal perception does not attribute properties to them.

known that our perception of the depicted scene remains relatively stable despite changes in viewing angle. This phenomenon was first analysed by Pirenne (1970), who suggested that the viewer compensates for the change in projection by taking into account the orientation of the picture surface. Nanay interprets this compensation as a function of dorsal processing, since it supports spatial interaction and egocentric localisation. He contrasts this with cases where such compensation fails, such as with distant ceiling frescoes, where dorsal access to the surface is impaired due to distance (Matthen 2005; Wollheim 1980; Polanyi 1970; Vishwanath et al. 2005).¹³ Then, Nanay argues that while the surface of the picture is always represented dorsally, it is not necessarily represented ventrally. That is, we can perceive a picture without attending to, or recognising, the picture surface. For example, when watching a football match on television, we perceive the depicted field without necessarily attending to the shape of the trapezoid projected on the screen. Clark (1960) discusses similar phenomena where surface properties must be actively noticed, rather than automatically processed. Nanay notes that this dissociation suggests that the ventral system is not necessarily involved in representing the surface.¹⁴ Finally, Nanay highlights empirical support for his overall dual-processing model by noting that damage to either the ventral or dorsal stream results in impaired picture perception. He reiterates that patients with ventral stream damage (visual agnosia) struggle with recognising depicted objects (Turnbull et al. 2004; Westwood et al. 2002). Conversely, patients with dorsal stream damage (optic ataxia) show specific deficits in assessing spatial relations in pictures.¹⁵ Taken together, these findings provide robust empirical support for Nanay's account of picture perception—and for the way our visual system functions when we see something in a picture. To sum up, his argument, which relies on the functional distinction between the ventral and dorsal visual subsystems, proposes that it is constitutive of picture perception that the ventral subsystem attributes properties to the depicted scene, while the dorsal subsystem attributes properties to the picture surface.¹⁶ This offers yet

¹³ Nanay also engages with experimental challenges to Pirenne's view. For example, Busey et al. (1990) showed that even in the absence of explicit orientation cues, subjects could still perceive faces in slanted pictures without distortion, raising doubts about whether surface orientation cues are necessary. Similarly, Halloran (1989) found cases in which distortion occurred despite the availability of surface cues. Nevertheless, Nanay defends a version of the compensation view by proposing a dorsal rather than ventral representation of picture surface orientation, which better explains the local and context-insensitive nature of this representation—consistent with the findings of Vishwanath et al. (2005) and the dorsal stream's general processing profile (Jeannerod 1997; Goodale and Milner 2004).

¹⁴ However, in cases where aesthetic appreciation is involved—such as when we attend to brushstrokes or composition—ventral representation of surface properties may come into play (Podro 1991, 1998; Lopes 2005; Hopkins 2010). This implies for Nanay that ventral processing of the surface is optional, not constitutive, for picture perception.

¹⁵ He cites the case of patient A.T., described by Jeannerod et al. (1994), who could perceive pictures but misjudged line lengths and object sizes—functions associated with dorsal processing. Although the empirical data on optic ataxia and picture perception are limited, Nanay notes that this partial dysfunction is exactly what his model would predict: damage to the dorsal stream does not eliminate picture perception but impairs specific aspects of it, namely spatial and surface-related judgments.

¹⁶ Nanay's claim that the dorsal stream does not represent the depicted scene appears to be in tension with some of the data presented by Briscoe (2016; see fn. 7). However, as Ferretti (2018) has argued, these data do not undermine the seeing-in theory. What is needed is a

another angle of support for the theory of seeing-in in particular, and for the development of an empirically informed theory of pictorial experience more broadly.

In this first section, I have shown, on the one hand, that the two empirical strands of Smith's methodology can indeed lend support to a key phenomenological claim—namely, that our experience of pictures is accurately described by the seeing-in theory. On the other hand, the discussion of picture perception also serves to reinforce the validity of Smith's methodological approach itself. While I hope that this contribution adds another dimension to Smith's project of developing a naturalized aesthetics of film—namely, a clearer understanding of the perceptual basis of our engagement with movies—in the next section I attempt to extend the discussion from low-level perceptual processes to the higher-order cognitive mechanisms that also seem necessarily involved in picture perception. Although vision science offers valuable insights into the nature of pictorial experience, it does not provide an account of depiction itself (Briscoe 2016). In fact, the psychological explanation of pictorial experience outlined in this section is not unique to pictorial artifacts; it also applies to the accidental images described by Wollheim in *Painting as an Art* (1987: 46–59) and cited by Smith—such as seeing faces and figures in clouds or on a stained wall. Pictures—whether still or moving—are artifacts intentionally created by someone, in a particular cultural context, usually for a communicative purpose. It is true, depictions are, first of all, based on the perceptual mechanism analysed in this section; without this initial level of perceptual engagement—without seeing-in—we could not even begin to speak about pictures and their representational specificity. Yet pictures are also based on intentional and cultural mechanisms—mechanisms related to conventions, norms, and rules.

3. Pictorial Experience and the Standard of Correctness

Smith acknowledges the problems faced by a hard-core perceptualist when he considers a possible skeptical worry raised by a hermeneutic theorist: “Appeals to science are all very well, the hermeneutic theorist might say, so long as we are only interested in the low-level, perceptual features of paintings and photos and movies. Vision scientists can tell us about edge detection and contrast effects, while their colleagues in the psychoacoustics lab can tell us about the McGurk, ventriloquist, and illusory flash effects. But they can't lay a glove on more complex meanings” (Smith 2017: 37).¹⁷ Smith affirms that one way to escape the charge of perceptualism is “to

more refined account of the possible interaction between the two visual streams in picture processing—an account that Ferretti himself provides, drawing on the most up-to-date empirical research in vision science.

¹⁷ Smith (2017: 37–38) suggests that there are essentially three main responses to this charge of perceptualism directed at naturalists like himself: (i) to emphasise that low-level features play a critical role; (ii) to stress that experiencing and appreciating artworks does not follow a simple one-way escalator from early vision to purely conceptual activity, but rather engages us in a cycle of perception, emotion, cognition, and reflection; and (iii) to point out that the character of many of the perceptual experiences generated by films and other artworks is often quite unlike ordinary perceptual experience. While I pursued line (i) in relation to pictures in the previous section, my aim here is to expand on (ii): to show that the experience of depiction does not follow a simple linear path from early vision to conceptual thought, but instead involves a cyclical process of perception, cognition, and reflection. Since Smith does not examine picture perception in detail, he does not develop these three

stress that experiencing and appreciating artworks does not follow a simple one-way escalator from early vision to purely conceptual activity, but rather engages us in a cycle of perception, emotion, cognition, and reflection” (Smith 2017: 37). For a naturalized aesthetics of film that also addresses the fundamental level of our visual engagement with movies as (moving) pictures—i.e. pictorial experience—this would mean exploring, for example, if, whether, and to what extent information provided by the author or by the context influences what (and how) we see in the picture. In this section, I propose a way in which a naturalized account could shed light on an aspect of the relationship between a subset of these norms—what Wollheim termed “the standard of correctness”—and pictorial experience.

For Wollheim, the standard of correctness is the set of non-visible information (regarding the production history of a picture) necessary for the correct viewing of an image. The standard of correctness has a normative role, it sets the norm by which we can “select the correct perception of a representation out of possible perceptions of it” (Wollheim 1980). In other words, the standard of correctness is a norm that sets the correct way of interpreting pictures. Wollheim held that this norm is provided by the intentions of the picture maker—it is because I, the author, didn’t intend to draw a tree in that spot of the picture’s surface that the latter doesn’t depict it. If it is the author that sets the correct interpretation of a picture, then the viewer can interpret the painting rightly or wrongly. Wollheim has a famous example to illustrate this: “In Holbein’s famous portrait in three quarters view (coll Thyssen) I normally see Henry VIII. However, I may have been going to too many old movies recently, and I look at the portrait and, instead of seeing Henry VIII, I now find myself seeing Charles Laughton” (Wollheim 1980: 206). The correct interpretation, Wollheim argues, is to see Henry VIII—as this was Holbein’s intention when painting the portrait.¹⁸ What Wollheim’s discussion highlights is that pictures are not merely perceptual; engaging with them also involves engaging with a normative layer that guides the selection of the correct interpretation. Wollheim was not entirely clear on whether the application of the standard of correctness genuinely enters into our pictorial experience—that is, whether it modifies the content of seeing-in.¹⁹ However, for a naturalized account of depiction, such as the one suggested by

lines of response specifically in relation to depiction and pictorial experience—even though, as we have seen, he does regard seeing-in as the basic form of expansionism. Part of (iii), therefore, was already addressed in the previous section. There are, however, other forms of expansionism that build on the peculiar perceptual nature of depictions and are not considered by Smith—for example, *inflection* (Podro 1991, 1998; Lopes 2005; Hopkins 2010) which also occur in the specific case of cinema (Hopkins 2020); perceptual illusions (such as motion-based illusions; see Marchetti 2022); non-linear perspectives; distortions; and specific filmic techniques such as montage or the *contra-zoom* (on the latter see Smith himself 2017: 39).

¹⁸ Newall (2011), building on Wollheim, argues that there are two different standards of correctness for two different types of images: handmade images—such as paintings, drawings, and traditional prints—have a standard of correctness established by the author’s intention; photographs and other images based on similar causal mechanisms have their standard of correctness set by a causal relationship (e.g. the photographic process).

¹⁹ Recently, the debate around the notion of pictorial experience has seen a resurgence of positions that, directly or indirectly, stem from Edmund Husserl’s theory of depiction—especially from his 1904/5 lectures *Phantasy and Image Consciousness* (Husserl 2005) and that seek to account for the *aboutness* of the image as a constitutive element of pictorial experience itself. In contrast to Wollheim’s twofoldness thesis, Husserl is taken to hold a

Smith's framework, this becomes a relevant and pressing question. I would like to conclude this section—and indeed this paper—by considering how one might begin to develop an answer to this question, both within Smith's framework and by extending it to include normativity.

The issue at stake is whether norms can shape perceptual experience in the specific case of picture perception. More precisely, in the context of this paper: does the standard of correctness affect the phenomenology of pictorial experience? While I do not attempt to argue for this claim here, I assume a positive answer—that is, that our overall experience changes when we follow the standard of correctness and adopt one interpretation of our perception over another—and, while a full answer would require a more detailed treatment, I conclude by briefly outlining two possible models of how such modification might operate—through cognitive penetration or cognitive phenomenology, understood as competing mechanisms—and finally suggest how Smith's methodology may offer a promising starting point for addressing the issue.²⁰ On one hand, cognitive penetration suggests that the application of concepts directly alters the phenomenology of perception, changing how viewers experience the image (Siegel 2011; Zeimbekis and Raftopoulos 2015; Stokes 2024). On the other hand, cognitive phenomenology proposes that the application of concepts supplements perception without fundamentally altering its character, instead adding a layer of normative understanding (Kriegel 2015). These two accounts point to different ways in which the standard of correctness might influence the viewer's engagement with the image. The choice between these mechanisms may depend on whether the viewer has prior familiarity with the depicted subject or pictorial conventions. For instance, experienced viewers might exhibit a more pronounced normative engagement, while novices may rely more heavily on immediate perceptual recognition. Importantly for this symposium, these issues could be addressed through both a more rigorous and systematic phenomenological analysis (of the kind suggested by Cavalletti in this symposium) and through an investigation into the underlying psychological mechanisms and neurophysiological structures—

threefold account of pictorial experience (see e.g. Kurg 2014). In fact, he describes the experience of seeing something in a picture—what he calls image consciousness—as involving three objects: the physical image [das physische Bild], i.e. the marked surface; the image object [Bildobjekt], i.e. what is seen in the surface (in this case, a man); i.e. the image subject [Bildsujet], i.e. what the picture is about (Charles Laughton or Henry VIII?). If, as it has been suggested (Eldridge 2018), Husserl takes image consciousness to be directed upon all such elements, then pictorial experience is not only twofold, but threefold. It is not clear though, the extent to which, if any, the consciousness of the third fold, the image subject, is perceptual for Husserl (see Voltolini 2018, 91; and Wiesing 2010). In fact, Nanay (2018, 175) reads Husserl's account as maintaining that the image subject is not perceived, but merely shows up in our judgment. The same goes, says Nanay, for a recent development of Husserl's theory, Lambert Wiesing's (2010) account. Indeed, for Wiesing, the image-carrier “displays” (2010: 35) the representation—or image-object which has a mere “artificial presence”—and is then related to the content, or image-subject. Appropriate image-subjects are those that resemble the image-object in certain respects (2010: 56). While the image-object is actually perceived, the image-subject is arguably non-perceptually apprehended. In this sense, however, the third fold plays the role that “the standard of correctness” plays in Wollheim's twofold account, hence threefoldness risks to boil down to a reformulation of twofoldness. See Marchetti 2024 for an overview. See Nanay 2018 and Voltolini 2018 for two recent competing accounts—*threefoldness* versus *three-layeredness*.

²⁰ While I only gesture toward this line of argument here, a full account will be developed elsewhere (see Marchetti and Terrone ms).

something that, as far as I know, has not yet been undertaken. A phenomenological approach would help clarify how normative expectations are experienced from a first-person perspective—for example, whether they are felt as interpretative constraints, background assumptions, or shifts in attention during picture viewing. At the same time, empirical research in cognitive psychology and neuroscience could shed light on how prior knowledge, conceptual frameworks, or cultural context influence our perceptual engagement with images, and whether these influences occur at early or late stages of processing. This combined approach, drawing on both philosophical and empirical resources, could lead to a better understanding of how perceptual and normative aspects interact in our experience of pictures. It would also allow us to explore whether and how normative considerations shape the perceptual content itself, not just its interpretation. For a naturalized aesthetics of film and pictorial art, such an approach would make it possible to explain the complexity of pictorial experience without losing sight of the fact that it is grounded in perceivable features of the image. If the standard of correctness does shape the phenomenology of picture perception, then understanding how this occurs—and under what conditions—is not only a philosophical question but also an empirical one. Investigating this could help clarify how normative elements integrate into a broader naturalistic account of depiction—one that takes seriously both the psychological mechanisms and the cultural context involved in picture perception. Further exploration of this issue, I believe, would contribute to the development of the kind of interdisciplinary aesthetics that Smith’s framework both presupposes and invites.²¹

References

- Anderson, B. and Winawer, J., 2005. Image segmentation and lightness perception. *Nature*, 434, 79–83.
- Berryhill, M.E. and Olson, I.R., 2009. The representation of object distance: evidence from neuroimaging and neuropsychology. *Frontiers in human neuroscience*, 3, 1–9.
- Briscoe, R., 2016. Depiction, pictorial experience, and vision science. *Philosophical topics*, 2, 43–81.
- Busey, T.A., Brady, N.P. and Cutting, J.E., 1990. Compensation is unnecessary for the perception of faces in slanted pictures. *Perception & psychophysics*, 48 (1), 1–11.
- Campbell, J., 2011. Sensorimotor knowledge and naïve realism. *Philosophy and phenomenological research*, 76, 666–673.
- Clark, A., 2001. Visual experience and motor action: are the bonds too tight? *Philosophical review*, 110, 495–519.
- Clark, K., 1960. *Looking at pictures*. London: John Murray.

²¹ I would like to thank Adriano d’Aloia, Federica Cavaletti, Murray Smith, Enrico Terzone and all other participants in the PEA Workshop “Film, Art, and the Third Culture”, held in Genoa on 13 December 2023, for their valuable feedback. This research was funded by the ERC-StG Project 101040535 (“PEA—The Philosophy of Experiential Artifacts”). Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

- Cutting, J., 2003. Reconceiving perceptual space. In: H. Hecht, R. Schwartz and M. Atherton, eds. *Looking into pictures: an interdisciplinary approach to pictorial space*. Cambridge, MA: MIT Press, 215–238.
- Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G. et al., 1998. Imaging unconscious semantic priming. *Nature*, 395, 597–600.
- Eldridge, P., 2018. Depicting and seeing-in: the 'sujet' in Husserl's phenomenology of images. *Phenomenology and cognitive sciences*, 17 (3), 555–578.
- Evans, G., 1982. *The varieties of reference*. Oxford: Oxford University Press.
- Evans, G., 1985. Molyneux's question. In: G. Evans, ed. *Collected papers*. Oxford: Clarendon, 364–399.
- Ferretti, G., 2018. The neural dynamics of seeing-in. *Erkenntnis*, 84 (6), 1285–1324.
- Ferretti, G., 2021a. Motoric understanding and aesthetic appreciation. *The journal for the philosophy of language, mind and the arts*, 2 (1), 113–130.
- Ferretti, G., 2021b. Why the pictorial needs the motoric. *Erkenntnis*, 88, 771–805.
- Fleming, R. and Anderson, B., 2004. The perceptual organization of depth. In: L. Chalupa and J. Werner, eds. *The visual neurosciences*. Cambridge, MA: MIT Press.
- Gilchrist, A., 1977. Perceived lightness depends on perceived spatial arrangement. *Science*, 195, 185–187.
- Gilchrist, A., 1980. When does perceived lightness depend on spatial arrangement? *Perception and psychophysics*, 28, 527–538.
- Gombrich, E.H., 1960. *Art and illusion*. New York: Pantheon Books.
- Gonzalez, C., Ganel, T., Whitwell, R., Morrissey, B. and Goodale, M., 2008. Practice makes perfect, but only with the right hand. *Neuropsychologia*, 46, 624–631.
- Goodale, M.A. and Humphrey, G.K., 1998. The objects of action and perception. *Cognition*, 67, 181–207.
- Goodale, M.A. and Milner, A.D., 1995. *The visual brain in action*. Oxford: Oxford University Press.
- Goodale, M.A. and Milner, A.D., 2004. *Sights unseen*. Oxford: Oxford University Press.
- Grill-Spector, K., Kourtzi, Z. and Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. *Vision research*, 41, 1409–1422.
- Halloran, T.O., 1989. Picture perception is array specific: viewing angle versus apparent orientation. *Perception & psychophysics*, 45, 467–482.
- Hopkins, R., 1998. *Picture, image and experience*. Cambridge: Cambridge University Press.
- Hopkins, R., 2008. What do we see in film? *Journal of aesthetics and art criticism*, 66 (2), 149–159.
- Hopkins, R., 2010. Inflected pictorial experience: its treatment and significance. In: C. Abell and K. Bantilaki, eds. *Philosophical perspectives on depiction*. Oxford: Oxford University Press, in press.
- Hopkins, R., 2020. Sculpting in time: temporally inflected experience of cinema. In: J. Pelletier and A. Voltolini, eds. *The pleasure of pictures: pictorial experience and aesthetic appreciation*. London: Routledge, 201–223.
- Husserl, E., 2005. *Phantasy, image consciousness, and memory*. Translated by J. Brough. Dordrecht: Springer.
- James, T., Humphrey, G., Gati, J., Menon, R. and Goodale, M., 2002. Differential effects of viewpoint on object-driven activation in dorsal and ventral stream. *Neuron*, 35, 793–801.

- Jeannerod, M., 1997. *The cognitive neuroscience of action*. Oxford: Blackwell.
- Jeannerod, M., Decety, J. and Michel, F., 1994. Impairment of grasping movements following a bilateral posterior parietal lesion. *Neuropsychologia*, 32, 369–380.
- Jeannerod, M., Decety, J. and Michel, F., 1994. Impairment of grasping movements following a bilateral posterior parietal lesion. *Neuropsychologia*, 32, 369–380.
- Kanwisher, N., 2004. The ventral visual object pathway in humans: evidence from fMRI. In: L. Chalupa and J. Werner, eds. *The visual neurosciences*. Cambridge, MA: MIT Press.
- Kriegel, U., 2015. *The varieties of consciousness*. New York: Oxford University Press.
- Kulvicki, J., 2014. *Images*. New York: Routledge.
- Kurg, R., 2014. Seeing-in as three-fold experience. *Postgraduate journal of aesthetics*, 11, 18–26.
- Lopes, D.M., 1997. Art media and the sense modalities: tactile pictures. *Philosophical quarterly*, 47, 425–440.
- Lopes, D.M., 2005. *Sight and sensibility*. Oxford: Oxford University Press.
- Marchetti, L., 2022. Depicting motion in a static image: philosophy, psychology and the perception of pictures. *British journal of aesthetics*, 62 (3), 353–371.
- Marchetti, L., 2024. Pictorial experience [online]. *International lexicon of aesthetics*. Spring 2024 edition. <https://lexicon.mimesisjournals.com/0Pictorial%20Experience.pdf> [Accessed 5 November 2025].
- Marchetti, L., 2025a. Seeing in mirrors, without seeing-in. *Philosophia*.
- Marchetti, L., 2025b. Seeing in VR, without seeing-in. *Estetika: the european journal of aesthetics*, LXII/XVIII (1), 71–81.
- Marchetti, L., 2025c. E pur si move! motion-based illusions, perception and depiction. *Australasian journal of philosophy*, 103 (3), 597–616.
- Marchetti, L. and Terrone, E., (ms). Pictorial experience and the standard of correctness.
- Marotta, J. and Goodale, M., 2001. The role of familiar size in the control of grasping. *Journal of cognitive neuroscience*, 13, 8–17.
- Matthen, M., 2005. *Seeing, doing, and knowing*. New York: Oxford University Press.
- Milner, A.D. and Goodale, M.A., 1995. *The visual brain in action*. Oxford: Oxford University Press.
- Nanay, B., 2011. Perceiving pictures. *Phenomenology and the cognitive sciences*, 10, 461–480.
- Nanay, B., 2018. Threefoldness. *Philosophical studies*, 175, 163–182.
- Nelissen, K., Joly, O., Durand, J.B., Todd, J.T., Vanduffel, W. and Orban, G.A., 2009. The extraction of depth structure from shading and texture in the macaque brain. *Plos one*, 4 (12), e8306.
- Newall, M., 2011. *What is a picture?* Basingstoke: Palgrave Macmillan.
- Newall, M., 2015. Is seeing-in a transparency effect? *British journal of aesthetics*, 55, 131–156.
- Niederée, R. and Heyer, D., 2003. The dual nature of picture perception. In: H. Hecht, R. Schwartz and M. Atherton, eds. *Looking into pictures*. Cambridge, MA: MIT Press, 77–98.
- Noë, A., 2002. On what we see. *Pacific philosophical quarterly*, 83, 57–80.
- Noë, A., 2003. Causation and perception: the puzzle unravelled. *Analysis*, 63, 93–100.
- Noë, A., 2004. *Action in perception*. Cambridge, MA: MIT Press.

- Peacocke, C., 1987. Depiction. *Philosophical review*, 96, 383–410.
- Pirenne, M.H., 1970. *Optics, painting, and photography*. Cambridge: Cambridge University Press.
- Podro, M., 1991. Depiction and the golden calf. In: N. Bryson, M.A. Holly and K. Moxey, eds. *Visual theory*. New York: Harper Collins, 163.
- Podro, M., 1998. *Depiction*. Cambridge, MA: Harvard University Press.
- Poincaré, H., 1958. *The value of science*. New York: Dover.
- Polanyi, M., 1970. What is a painting? *British journal of aesthetics*, 10, 225–236.
- Rice, N.J., Valyear, K.F., Goodale, M.A., Milner, A.D. and Culham, J.C., 2007. Orientation sensitivity to graspable objects: an fMRI adaptation study. *Neuroimage*, 36, T87–T93.
- Sakata, H., Tsutsui, K. and Taira, M., 2003. Representation of the 3D world in art and in the brain. *International congress series*, 1250, 5–35.
- Siegel, S., 2011. Cognitive penetrability and perceptual justification. *Noûs*, 46 (2), 201–222.
- Smith, M., 2017. *Film, art, and the third culture: a naturalized aesthetics of film*. Oxford: Oxford University Press.
- Stokes, D., 2021. *Thinking and perceiving: on the malleability of the mind*. New Problems of Philosophy series. London: Routledge.
- Taira, M., Nose, I., Inoue, K. and Tsutsui, K., 2001. Cortical areas related to attention to 3D surface structures based on shading: an fMRI study. *Neuroimage*, 14, 959–966.
- Terrone, E., 2024. Are works of art affective artifacts? If not, what sort of artifacts are they? *Topoi*, 43 (3), 737–746.
- Tsutsui, K., Sakata, H., Naganuma, T. and Taira, M., 2002. Neural correlates for perception of 3D surface orientation from texture gradient. *Science*, 298, 409–412.
- Tsutsui, K., Taira, M. and Sakata, H., 2005. Neural mechanisms of three-dimensional vision. *Neuroscience research*, 51, 221–229.
- Turnbull, O.H., Driver, J. and McCarthy, R.A., 2004. 2D but not 3D: pictorial depth deficits in a case of visual agnosia. *Cortex*, 40, 723–738.
- Ullman, S., 1979. The interpretation of structure from motion. *Proceedings of the royal society of London b: biological sciences*, 203, 405–426.
- Vishwanath, D., Girshick, A.R. and Banks, M.S., 2005. Why pictures look right when viewed from the wrong place. *Nature neuroscience*, 8, 1401–1410.
- Voltolini, A., 2003. In defense of seeing-in. In: H. Hecht, R. Schwartz and M. Atherton, eds. *Looking into pictures*. Cambridge, MA: MIT Press, 3–15.
- Voltolini, A., 2013. Defiction. In: C. Barbero, M. Ferraris and A. Voltolini, eds. *From fictionalism to realism*. Cambridge: Cambridge Scholars Publishing, 43–61.
- Voltolini, A., 2015. *A syncretistic theory of depiction*. London-New York: Palgrave Macmillan.
- Voltolini, A., 2018. Twofoldness and three-layeredness in pictorial representation. *Estetika: the central European journal of aesthetics*, 55 (1), 89–111.
- Wallach, H. and O'Connell, D.N., 1953. The kinetic depth effect. *Journal of experimental psychology*, 45 (4), 205.
- Walton, K., 1990. *Mimesis as make-believe*. Cambridge, MA: Harvard University Press.

- Westwood, D.A., Danckert, J., Servos, P. and Goodale, M., 2002. Grasping two-dimensional images and three-dimensional objects in visual-form agnosia. *Experimental brain research*, 144, 1432–1106.
- Wiesing, L., 2010. *Artificial presence*. Translated by N.F. Schott. Stanford, CA: Stanford University Press.
- Wollheim, R., 1980. *Art and its object*. 2nd ed. Cambridge: Cambridge University Press.
- Wollheim, R., 1987. *Painting as an art*. Princeton: Princeton University Press.
- Zeimbekis, J., 2015. Seeing, visualizing, and believing: pictures and cognitive penetration. In: J. Zeimbekis and A. Raftopoulos, eds. *The cognitive penetrability of perception: new philosophical perspectives*. Oxford: Oxford University Press, 298–327.
- Zeimbekis, J. and Raftopoulos, A., 2015. *The cognitive penetrability of perception: new philosophical perspectives*. Oxford: Oxford University Press.