In Michael Beaney (ed.), Aspect Perception After Wittgenstein: Seeing-as and Novelty. Routledge. pp. 49-88 (2018)

3 Gombrich and the Duck–Rabbit

Robert Briscoe

> An oil painting caught and held him. . . . There was beauty, and it drew him irresistibly. He forgot his awkward walk and came closer to the painting, very close. The beauty faded out of the canvas. His face expressed his bepuzzlement. He stared at what seemed a careless daub of paint, then stepped away. Immediately all the beauty flashed back into the canvas. 'A trick picture', was his thought . . .
>
> —Jack London, *Martin Eden* (1909/1982)

1. Introduction

A picture is a marked or otherwise patterned, two-dimensional (2D) surface that, when present to sight, elicits the experience as of an absent, three-dimensionally (3D) organized scene.[1] How ought we to understand the nature of this experience? There are a number of distinct suggestions in the philosophical literature:

A. Pictures elicit a 3D-scene-representing experience of the same psychological kind as the experience of seeing face-to-face (Gombrich 1961/2000, 1972, 1982; Briscoe 2016). Since the represented, 3D scene is absent from sight, however, this experience is non-veridical.
B. When we look at a picture, we enjoy an experience as of the depicted, 3D scene. This experience, however, is always fused with awareness of the superficial pattern on the pictorial surface (Wollheim 1987, 1998, 2003). In this respect, pictorial experience has two different dimensions or 'folds' of representational content.

Accounts (A) and (B) maintain that the appearance of depth and 3D structure 'beyond' the 2D pictorial surface is essential to pictorial experience. When we look at a suitably patterned surface, Richard Wollheim writes, we are typically aware of 'something in front of or behind something else' (1998: 221). As John Kulvicki puts it, 'there is a strong sense in which depicted scenes seem to recede from the canvas' (2009: 391). Other prominent accounts, by contrast, attach theoretical priority to seeing properties

of the pattern or design visible *on* the pictorial surface. They agree with Malcolm Budd that

> the way to capture the experience of seeing a picture as a depiction of its subject is not by adding any other visual experience to the visual awareness of the picture-surface—either as a separate experience or by fusing it with the visual awareness of the picture-surface—but by specifying the nature of the visual awareness of the picture-surface when you see what the picture depicts.
>
> (1992/2008: 203–204)

There are two familiar approaches to characterizing the nature of this surface-awareness:

C. To experience a 3D scene in a picture's surface is to see the latter as resembling the former in certain respects (Budd 1992/2008, 1993/2008; Hopkins 1998, 2006). For example, to experience a cubical object oriented in depth when looking at a drawing of the Necker cube is to see the pattern of lines that make up the drawing as resembling such an object.
D. Pictures function as props in 'visual games of make-believe' (Walton 1990, 2008). In looking at a picture, the viewer imagines of her experience that it is a seeing of whatever the picture portrays. 'In the case of picture perception, not only does looking at the picture induce us to imagine seeing an ox, we also imagine our actual visual experience, our perceiving the relevant part of the canvas, to be an experience of seeing an ox' (Walton 2008: 118).[2]

In the philosophy of art, E.H. Gombrich is by far the best-known proponent of option (A), according to which pictorial experience is psychologically continuous with the experience of seeing face-to-face. Pictures, Gombrich says, have the power to arouse in us a 'visual experience of a kind that we know from our encounters with reality' (1982: 181). In what follows, I will refer to this view as the *Continuity Hypothesis*.

The Continuity Hypothesis comprises three main claims. First, a picture is a patterned, 2D surface designed to elicit a non-veridical experience as of depth and 3D structure—'something akin', as Gombrich puts it, 'to a visual hallucination' (1972: 208). Crucially, this hallucination-like experience represents its intentional objects as located on the far side, as it were, of the patterned, pictorial surface. Gombrich, as we have already seen, is not alone in characterizing the spatial phenomenal character of pictorial experience in this way. 'The first intention of the painter', Leonardo da Vinci tells us in his *Treatise on Painting*, 'is to make a flat surface display a body as if modeled and separate from this plane' (quoted by Kemp 1989: 15). A picture, as J.J. Gibson puts it, 'is both a scene and a surface, and the scene is paradoxically *behind* the surface' (1979: 281). When we look at *The Peasant Wedding* (Figure 3.1), for example, we do not merely perceive an array of colours

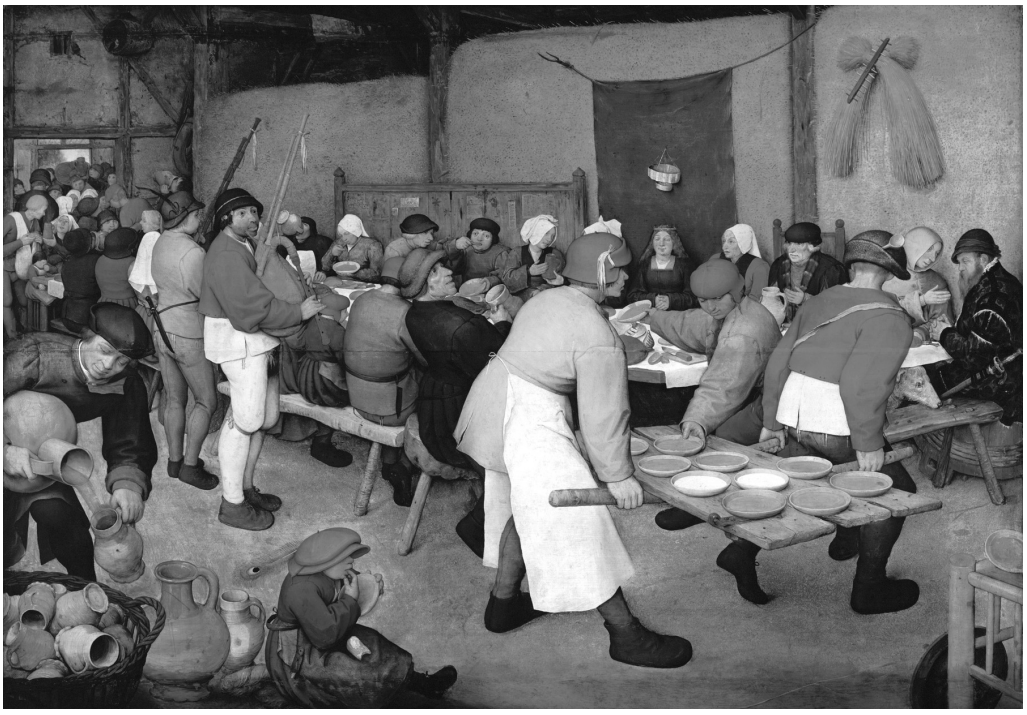

Figure 3.1 Pieter Bruegel the Elder, *The Peasant Wedding* (1567)
Oil on panel, 124 cm × 164 cm, Vienna, Kunsthistorisches Museum.

located on a single plane of depth. We also experience a complexly organized, 3D scene in which voluminous objects participate in a 'recessional movement', to borrow Wölfflin's (1929) phrase, from the pictorial point of view. Following standard usage in art history, aesthetics, and perceptual psychology, I will refer to the virtual, 3D space in which we visually experience shapes, sizes, colours, textures, orientations, and other features when we look at a picture as *pictorial space* (Wölfflin 1929; White 1967; Pirenne 1970; Kubovy 1986; Rogers 1995, 2003; Koenderink 1998, 2012; Hecht *et al*. 2003; Thompson *et al*. 2011: ch. 12).

The second claim is that phenomenological and representational continuities between pictorial experience and seeing face-to-face reflect their underlying, psychological continuity. Both experiences result, Gombrich says, from an unconscious, inference-like process of 'guided projection'.

On the one hand, the process is said to be *guided* because it relies on certain nonconscious expectations and assumptions about the structure of the visible environment in order to interpret the message conveyed by the retinal image. These assumptions perform two closely related functions: they are used by the perceptual system to generate a 'hypothesis' about the most probable cause of the image in the distal environment and, in addition, to test that hypothesis by generating predictions concerning, among other things, the sensory effects of bodily movement: 'every message sets up a set

of expectations with which the incoming flow can be matched to confirm correct assessments or to modify and knock out false guesses' (Gombrich 1978: 158; see also 1961/2000: 274–275). Gombrich refers to the contribution of these background assumptions to perception as the 'beholder's share'.

On the other hand, the process is said to involve *projection* because its conscious end product is underdetermined by the structure of the light sampled by the eye. A given retinal image could be caused by the light reflected from (or emitted by) many different 3D scenes. (This is the so-called inverse optics problem in vision science.) Contemporary Bayesian 'predictive coding' models of visual processing replace talk of implicit assumptions with talk of prior probabilities and likelihoods (for discussion, see Hohwy 2013; Clark 2015; Rescorla 2015), but in other key respects, they bear strong affinities to the Gombrichian account. On both approaches,

> perception is a process . . . in which we (or rather, various parts of our brains), try to guess what is out there, using the incoming sensory signal more as a means of tuning and nuancing the guessing rather than as a rich . . . encoding of the state of the world.
>
> (Clark 2015: 27)

The following passage from *Art and Illusion* so perfectly captures the kinship between the Gombrichian account of perception and the more recent predictive coding framework that it is worth quoting at length:

> The experience of the radio 'monitor' confronted with indistinct speech and that of the sailor confronted with indistinct shapes on the horizon are not incommensurate. We must always rely on guesses, on the assessment of probabilities, and on subsequent tests, and in this there is an even transition from the reading of the symbolic material to our reaction in real life. When we wait at the bus stop and hope the Number Two is coming into sight, we probe the indistinct blot that appears in the distance for the possibility of projecting the number 'two' into it. When we are successful in this projection, we say we now see the number. This is a case of symbol reading. But is it different with the bus itself? Certainly not on a foggy night. Nor even in full daylight, if the distance is sufficiently great. Every time we scan the distance we somehow compare our expectation, our projection, with the incoming message. If we are too keyed up, as is well known, the slightest stimulus will produce an illusion. Here as always it remains our task to keep our guesses flexible, to revise them if reality appears to contradict, and to try again for a hypothesis that might fit the data. But it is always we who send out these tentacles into the world around us, who grope and probe, ready to withdraw our feelers for a new test.
>
> (Gombrich 1961/2000: 178–179)

Like the predictive coding approach, the Gombrichian account of perception turns a traditional, input-dominated view of how vision works on its head. Visual experience is not a passive imprinting of the world on the mind but, rather, an active, hypothesis-generating, and hypothesis-testing perceptual process: 'The world never presents a neutral picture to us; to become aware of it means to become aware of possible situations that we can try out to test for their validity' (1961/2000: 275).

The central role of prediction and projection in perception is masked, Gombrich suggests, by the rapid, automatic, and seemingly effortless way in which the 3D structure of a visible scene is usually revealed to us. This point is illustrated in a discussion of a natural history engraving of some plants, insects, and animals displayed on an unstructured, white background:

> Looking at Jacob Hoefnagel's plate . . . , we always supply the appropriate ground to the figure: the lizard sits on a slope, while some insects, throwing shadows, are imagined against a flat ground and others are seen as flying. Without knowing it, we have carried out a rapid succession of tests for consistency and settled on those readings which make sense.
>
> (1961/2000: 231)

The pictorial arts, as this passage makes clear, provide an especially useful arena for probing the role of the beholder's share in perception. The information in the light available to the eye, as Gombrich emphasizes, is 'immeasurably richer . . . when we move around in the real world' (1961/2000: 274) than when viewing an etching, drawing, or painting of a 3D scene. In consequence, pictures afford special opportunities for teasing out the relative contributions of 'top-down' projection and the 'bottom-up' sensory signal to perceptual processing. The more ambiguous or incomplete the signal from the environment, the more prominent the role played by the beholder's share in the process of visual hypothesis formation. 'The deliberately blurred image, the *sfumato*, or veiled form . . .', for example, 'cuts down the information on a canvas and thereby stimulates the mechanism of projection' (1961/2000: 175–176).

Finally, it is central to Gombrich's project in *Art and Illusion* that the ambiguity of the environment's image on the retina 'can never be seen as such' (1961/2000: 249). The visual system selects only one consistent, 3D-scene interpretation of the image at a time, even when more than one such interpretation can be made to 'fit'.[3] In support of this claim, Gombrich appeals to ambiguous or 'multistable' figures in which figure-ground assignments, shapes, orientations, groupings, or other organizational properties appear to alternate with prolonged inspection. Whether side *abcd* or side *efgh* appears closer in depth varies from one moment to the next, when we look at a drawing of the Necker cube (Figure 3.2), but we never experience both organizations at once. The neuroscientists David Leopold and

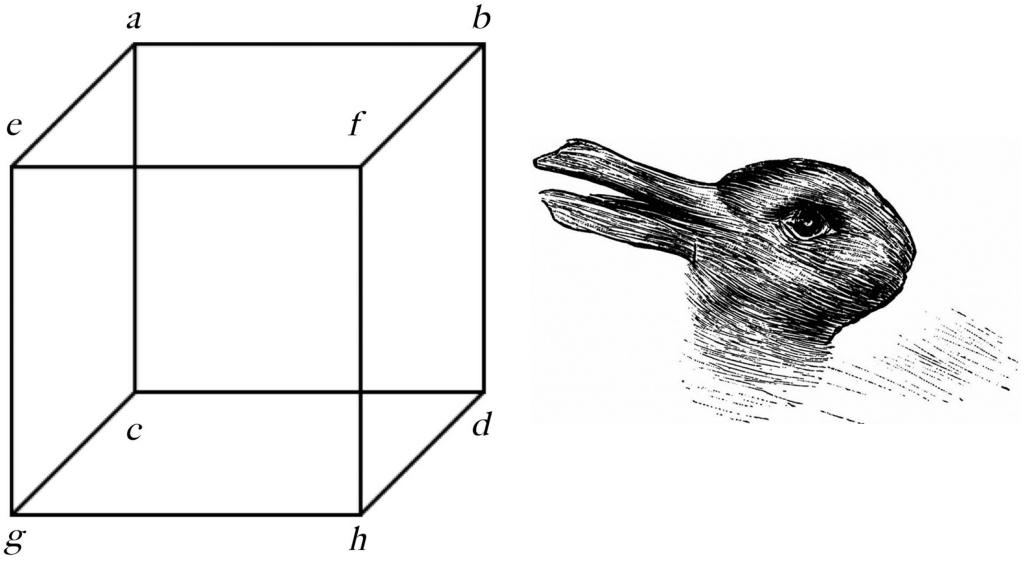

Figure 3.2 Ambiguous figures: the Necker cube and the duck–rabbit

Nikos Logothetis refer to this familiar property of multistable perception as 'exclusivity':

> Exclusivity, or uniqueness, ensures that conflicting visual representations are never simultaneously present [in conscious awareness]. That only a single perceptual solution can exist at once is likely to have its origins in the structure of the sensory machinery itself; that is, uniqueness is a fundamental encoding principle among neurons in the visual cortex.
>
> (Leopold and Logothetis 1999: 260)

Gombrich argues that the exclusivity of visual processing—the requirement of a single, consistent, 3D scene interpretation—has important implications for understanding the nature of pictorial experience. When an observer looks at *The Peasant Wedding* (Figure 3.1), for example, two conflicting sets of depth cues are typically available to her visual system. At close range, binocular disparity, convergence, and accommodation specify the orientation of the canvas and its distance in depth. The structured array of light reflected from the painting's surface, however, is also a vehicle for sources of spatial information that jointly specify the layout and properties of objects in a (virtual) three-dimensionally organized scene. These monocular or 'pictorial' depth cues, as they are sometimes called, include, but are not limited to, occlusion, texture gradients, shadows, reflections, relative size, linear perspective, atmospheric haze, height in the visual field, and the horizon ratio (for useful reviews, see Ames 1925; Cutting and Vishton 1995; Palmer 1999; and Thompson *et al.* 2011). In consequence, a picture may elicit either

of two very different experiences: an experience that attributes properties to the patterned, 2D pictorial surface *or* an experience that attributes properties to objects in phenomenally 3D pictorial space. We can look at *The Peasant Wedding*, or we can look *into* it.

Gombrich argues that while it is possible to alternate or 'switch' between these different experiences, it isn't possible to enjoy both of them at once:

> But is it possible to 'see' both the plane surface and the battle horse at the same time? If we have been right so far, the demand is for the impossible. To understand the battle horse is for a moment to disregard the plane surface. We cannot have it both ways.
>
> (1961/2000: 279)

In this respect, Gombrich suggests, there is an analogy with the experience of looking at an ambiguous figure. Just as noticing the duck aspect in a drawing of the duck–rabbit (Figure 3.2) excludes noticing the rabbit aspect at the same time, visually experiencing the pattern on an opaque, pictorial surface excludes visually experiencing the way objects are arranged 'beyond' that surface in pictorial space. In both cases, Gombrich suggests, 'We are not aware of the ambiguity as such, but only the various interpretations. . . . We can train ourselves to switch more rapidly, indeed to oscillate between readings, but we cannot hold conflicting interpretations' (1961/2000: 236).

The now standard objection to this analogy is that while the two interpretations in the first case (duck vs rabbit) do indeed genuinely conflict, the two interpretations in the second case (2D surface vs 3D scene) are merely *different*. Richard Wollheim complains:

> But by what right does Gombrich assume that we can no more see a picture as canvas and as nature, than we can see the duck-rabbit figure as a duck and as a rabbit? Because—it might be said—canvas and nature are different interpretations. But if this is Gombrich's argument, it is clearly invalid. For we cannot see the duck-rabbit figure as duck and as rabbit, not because these are two different interpretations, but because they are two incompatible interpretations.
>
> (Wollheim 1963: 29)

A similar objection has been voiced by Dominic Lopes:

> the duck–rabbit figure undermines Gombrich's intended use of it. . . . Switches between the two contents (as of duck and rabbit respectively) are not analogous to switches between the figure's design, on the one hand, and either of its contents, on the other. That duck cannot be seen simultaneously with rabbit fails to show that duck cannot be seen simultaneously with design or design simultaneously with rabbit.
>
> (Lopes 2005: 31)

My primary purpose in the remainder of this chapter is to clarify, develop, and defend the Gombrichian account of pictorial experience, drawing on resources from contemporary vision science. Section 2 provides motivation for Gombrich's use of the duck–rabbit analogy by situating it in the context of his attack on a historically influential conception of the appearance–reality distinction in visual perception. According to the conception in question, everyday visual experience comprises two distinct 'layers' or 'folds' of representational content. The content of the first layer represents an object's viewpoint-dependent appearance or 'look', where this is supposed to be phenomenally flat or 2D in character. The content of the second layer, by contrast, is supposed to represent the object's intrinsic, viewpoint-independent properties as well as its distance in depth. Pictorial perception is arguably the best case for this 'dual content' theory of visual experience, since a picture really is a flat patchwork of colours that, when present to sight, elicits the impression of depth and 3D structure. One of Gombrich's main aims in *Art and Illusion* is to demonstrate that the theory fails even as an account of pictorial experience.

The dual content theory of pictorial experience is seemingly paradoxical. It maintains that when we look at a picture, the same solid angle in the visual field is represented *twice over*—once as filled by an opaque, two-dimensionally organized surface and once as containing a three-dimensionally organized scene. **Section 3** argues that two of the better-known attempts to eliminate the appearance of paradox at the heart of the dual content theory of pictorial experience are unsuccessful. The two putative layers of representational content in pictorial experience, it concludes, contrary to Wollheim, aren't merely different. As Gombrich insists, they are also incompatible.

Section 4 defends and constructively elaborates on the Continuity Hypothesis. In particular, it argues for an account of the structure of pictorial experience that I refer to as *weak onefoldness*. Pictorial experience is onefold in the sense that its content reflects a single, consistent, 3D-scene interpretation of the retinal image. Pictorial experience is only weakly onefold, however, in that it typically attributes certain combinations of properties to the 2D pictorial surface and to objects in phenomenally 3D pictorial space at the same time. Having the experience of virtual depth and 3D structure, when looking at a picture, I argue, excludes representing some, but not *all*, of a picture's surface properties.

A second aim of section 4 is to reconcile the claim that pictorial experience and seeing face-to-face are psychologically continuous with the observation that the former experience does not typically dispose the perceiver to believe that its objects are real. A recent account of *stereopsis* from Dhanraj Vishwanath (2014), I propose, makes such reconciliation possible. According to Vishwanath, pictorial experience does not dispose the perceiver to believe that its objects are present to sight because, in contrast with ordinary, non-pictorial visual experience, it fails to specify their locations at certain absolutely scaled distances in depth. This contrast, I argue, however,

is best understood as a difference at the level of representational content rather than a difference at the level of psychological kind.

2. Pictorial Experience and the Denial of Dual Awareness

Gombrich, as we have seen, denies that it is psychologically possible to experience the properties of the pattern visible *on* a 2D pictorial surface, while simultaneously experiencing depth and 3D structure *in* that surface. These two interpretations or 'readings' both fit the configuration of light intensities on the retina but are mutually exclusive. This view is central to argument of *Art and Illusion*. Gombrich goes so far as to identify our inability to experience the ambiguity of the incoming retinal signal as the 'theme song' of the book (1961/2000: 313).

Gombrich, it is important to emphasize, doesn't just reject a dual content theory of pictorial experience. He also rejects of a dual content theory of ordinary, non-pictorial seeing. The primary target of Gombrich's criticism in *Art and Illusion* is, in fact, a historically influential conception of the appearance-reality distinction in everyday visual perception and its application to the special case of perceiving pictures.

According to the conception in question, the representational content of visual experience divides into two 'layers'. The first, putative layer of visual content represents the viewpoint-dependent appearance or 'look' of a perceived object, where this is supposed to be phenomenally 2D in character, much like a flat projection of the object's shape on the frontal plane. In addition, the first layer of content is frequently alleged to be phenomenologically, epistemically, and/or developmentally more basic than the second layer.[4] 'What we really see', according to one prominent version of the dual content theory, 'is a medley of colored patches such as Turner paints' (Gombrich 1961/2000: 296). The 19th-century art critic John Ruskin provides an especially clear expression of this outlook in *The Elements of Drawing*:

> The perception of solid Form is entirely a matter of experience. We *see* nothing but flat colours. . . . The whole technical power of painting depends on our recovery of what might be called the innocence of the eye; that is to say, of a sort of childish perception of these flat stains of colour, merely as such, without consciousness of what they signify—as a blind man would see them if suddenly gifted with sight.
>
> (Ruskin 1856/1971: 27)[5]

The second, putative layer of visual content, by contrast, is supposed to represent an object's intrinsic, viewpoint-independent properties as well as its distance and orientation in depth. It results from the application of perceptual constancy mechanisms, spatial organizational principles, and learned associations to the content of the first layer. Gombrich variously associates versions of this dual content theory of visual experience with the British

Empiricist tradition in philosophy, the introspectionist movement in psychology, and Impressionism's 'discovery of appearances' (Fry 1934).

The dual content of theory of visual experience is not just a historical curiosity. Contemporary advocates in philosophy include William Lycan (1996, 2008), Jonathan Cohen (2010), and Berit Brogaard (2012). Perhaps the most influential recent proponent of the theory, however, is Alva Noë (2004, 2005):

> Perceptual constancy—size and shape constancy—coexists with perspectival *nonconstancy*. Two tomatoes, at different distances from us, may visibly differ in their apparent size even as we plainly see their sameness of size; a silver dollar may look elliptical—when we view it from an angle, or when it is tilted in respect of us—even though it also looks, plainly, circular.
>
> Perceptual experience presents us with the world (the constancies) and it presents us with how the world perceptually seems to be (the nonconstancies). A satisfying account of perception must explain how the silver dollar can look both circular and elliptical, how the tomatoes can look to be the same in size and yet different in size. Perceptual experience is two-dimensional, and this needs explaining.
>
> (Noë 2005: 235)

Noë identifies an object's non-constant, visually apparent shape with the shape of the patch that would perfectly occlude the object on a plane perpendicular to the line of sight. He refers to this as the object's 'perspectival shape' (P-shape). An object's non-constant, 'perspectival size' (P-size), in turn, corresponds to the size of the patch that would occlude the object on the same plane. Non-constant P-properties are 'perceptually basic' (2004: 81), according to Noë, because in order to see an object's constant spatial properties it is necessary both to experience its P-properties and to understand how these would undergo transformation across changes in one's point of view. Seeing is thus a 'two-step' process: 'How they (merely) appear to be plus sensorimotor knowledge gives you things as they are' (Noë 2004: 164).

Although we typically attend to the content of the second, post-constancy layer in our everyday interactions, the content of the first layer, according to dual content theorists, is supposed to remain introspectively accessible, if only we know how to look for it. 'There is a sense . . .', Noë writes, 'in which we move about in a sea of perspectival properties and we are aware of them (usually without thought or notice) whenever we are perceptually conscious. Indeed, to be perceptually conscious is to be aware of them' (Noë 2004: 167). We normally attend to the circular shape of an obliquely viewed coin, but this doesn't mean that it ceases to look elliptical when we do. As paradoxical as it may sound, a silver dollar can look circular and elliptical at the same time.

The Gestalt psychologists referred to the assumption that pre-constancy visual appearances co-exist alongside our post-constancy representations of the world as the doctrine of 'unnoticed sensations' (Köhler 1913/1971). In *The Ecological Approach to Visual Perception*, J. J. Gibson characterizes this doctrine in following way:

> It has been generally believed that even adults can become conscious of their visual sensations if they try. You have to take an introspective attitude, or analyze your experience into its elements, or pay attention to the data of your perception, or stare at something persistently until the meaning fades away. I once believed it myself. I suggested that the 'visual field' could be attended to, as distinguished from the 'visual world', and that it was almost a flat patchwork of colors, like a painting on a plane surface facing the eye . . .
>
> (Gibson 1979: 286)

Using Gibson's terminology, the first layer of visual representational content posited by the dual content theory represents the structure of the two-dimensionally organized visual field, while the content of the second layer represents the layout of the three-dimensionally organized visual world.

Gombrich, like Gibson, is deeply sceptical of the doctrine of unnoticed sensations. Perceptual constancy mechanisms, he argues, do not operate on a platform of conscious, but normally inconspicuous, 2D appearances or P-properties. The first, introspectively accessible product of sub-personal, perceptual information processing is an experience that represents the disposition and properties of objects in the 3D visual world. Gombrich writes:

> Presented with a circular disk, for instance, we are well aware of the fact that it might be fairly large and far away, or small and close by. We also may remember intellectually that it might be a tilted ellipse, or a number of other shapes, but we cannot possibly see these infinite possibilities; the disk will appear to us as an object out there, even though we may realize, as students of perception, that another person may guess differently.
>
> One must have experienced these effects to realize how elusive they make the idea of 'appearance' as distinct from the object itself. The stimulus school of psychology and the phenomenalists talked as if the 'appearance' of the disk, the stimulus pattern, were the only thing really 'experienced' while all the rest was inference, interpretation. It sounds like a plausible description of vision, but it is untrue to our actual experience. We do not observe the appearance of color patches and then proceed to interpret their meaning. . . . To see is to see 'something out there'.
>
> (Gombrich 1961/2000: 260)

> It is even harder to see the world as a two-dimensional field than it is to see one's own image on the mirror's surface. Our belief that we can

> ever make the world dissolve into such a flat patchwork of colors rests itself on an illusion . . .
>
> It is to the three-dimensional world that our organism is attuned, where it learns to test its anticipations against the flow of incoming stimuli. . . .
>
> (Gombrich 1961/2000: 328–329)

Ruskin's 'flat stains of color' and Noë's P-properties, on this view, are not a foundational, pre-constancy layer of conscious experience on which a further layer of post-constancy, visual representational content is constructed. Instead, perceptual processing begins with texture gradients, edge junctions, and other properties of the retinal image that are predictive of properties in the environment and, on this nonconscious evidential basis, constructs a representation of the most probable, three-dimensionally organized scene—a 'possible configuration in space and light' (Gombrich 1961/2000: 327). Conscious visual experience, in other words, begins only after perceptual constancy mechanisms and Gestalt organizational principles have done their work, that is, with the Gibsonian visual world as opposed to the Gibsonian visual field (Briscoe 2008). The eye is never innocent in Ruskin's sense.

The Gombrichian account, as already observed, has a close counterpart in contemporary vision science. The central challenge faced by the visual system, according to recent Bayesian models of perception (for overviews, see Knill and Richards 1996; Mamassian *et al*. 2002; Clark 2013, 2015; Hohwy 2013; Rescorla 2015), is to infer the most probable cause of the retinal image on the basis of two sources of 'evidence': (1) the various depth cues present in the image itself, as well as (2) learned or innate assumptions about the statistical properties of the natural environment and the image formation process. The content of the perceptual state formed in response to a particular pattern of retinal stimulation—the brain's operative 'hypothesis' about the structure of the impinging environment—is the cause to which the highest probability is assigned given all the available endogenous and exogenous evidence. In general, this will be one of many different possible three-dimensionally organized scenes: the *hypothesis space* for causal inference in vision is a 3D-scene space, in which different hypotheses correspond to different possible arrays of objects at a distance from the perceiver's eyes.[6] (One such hypothesis picks out the very scene in front of the reader now.) Crucially, each interpretation of the retinal image contained in the hypothesis space for vision is a *post-constancy* interpretation. Interpretations in the hypothesis space for vision, this is to say, range over possible states of the Gibsonian visual world rather than Gibsonian visual field.[7] To see is always to see, as Gombrich says, 'something out there'.

Neither Gombrich nor Gibson would deny, of course, that there are contexts in which an observer will experience an obliquely viewed disk as an ellipse in the frontal plane. When the disk is poorly illuminated, or far away, or seen under experimentally contrived 'reduced cue' conditions, such

non-veridical perception is entirely possible. It seems clear, however, that the possibility of visual illusion under informationally impoverished viewing conditions by itself provides no support for the dual content theorist's claims about the way the world appears to us under normal, informationally rich viewing conditions (Briscoe 2008; Hopp 2013). No more, say, than the possibility of mistaking Sarah Palin for Hillary Clinton on a dark night provides support for the claim that Palin looks like Clinton in broad daylight.

For present purposes, there are three important points. First, that a disk may look elliptical in the contexts mentioned in the last paragraph does no work when it comes to motivating the claim that there is a conscious, pre-constancy layer of perceptual representational content. On the contrary, the non-veridical, *ellipse-in-the-frontal-plane* interpretation of the retinal image *is* a post-constancy interpretation, in particular, the post-constancy interpretation that is most probable in light of the (meagre) information available to the visual system. The disk appears to be intrinsically elliptical in shape—full stop. There is just one layer of post-constancy, visual representational content.

Second, the two interpretations, viz., *disk-slanted-in-depth* and *ellipse-in-the-frontal-plane*, are clearly incompatible. A single, opaque surface cannot look to be intrinsically circular and slanted in depth and intrinsically elliptical and at a right angle to the line of sight at the same time (or at different times, holding sources of optical information and viewing conditions fixed).

Finally, far from having a developmentally, phenomenologically, or epistemically privileged status, the elliptical appearance, in this context, is simply evidence of a breakdown in perceptual constancy. Veridical shape perception is not guaranteed when the visual system operates outside of information-rich, ecologically normal viewing conditions. The real contrast between the two interpretations of the retinal image, in short, is not that only one of them is 'innocent' but, rather, that only one of them also is accurate.

Recent empirical work on the role of depth in perceptual organization is helpful for purposes of further elucidating Gombrich's view. Roland Fleming and Barton Anderson (2004) divide 'legal interpretations' of a luminance edge in the retinal image into two main classes. Interpretations in the first class represent a *surface event* in which both sides x and y of the edge are located at the same distance in depth from the perceiver. Examples include reflectance edges, cast shadows, and creases on a single, opaque surface. Interpretations in the second class represent the contour of a background-occluding object and, hence, a difference in depth at the edge: one side of the edge, either x or y, is located at the depth of the object while the other side of the edge lies at the more distant depth of the partially occluded background (Fleming and Anderson 2004: 1287). In this case, the edge is said to be 'owned' by the object (Nakayama *et al.* 1995).

In terms of this framework, Gombrich's view is that, when looking at a picture, the visual system cannot legally interpret a given edge or set of edges in the retinal image in both ways at the same time. What the visual system

can do, he suggests, is alternate or switch between a 'surface event' interpretation of an edge and an 'object contour' interpretation of the edge. These different interpretations do not correspond to different, simultaneously accessible layers of content in one's experience of the picture but, rather, to two, temporally distinct experiences.

An analogous Gombrichian claim holds for how we experience non-spatial properties, such as lightness (albedo) and colour, when looking at a picture. Consider the lightness illusion by Barton Anderson and Jon Winawer (2005) reproduced in Figure 3.3. In the figure, the disks on the light and dark surrounds are photometrically identical, but the disks on the light surround appear as uniformly black objects visible behind a semi-transparent, light haze, whereas the disks on the dark surround appear as uniformly white objects visible behind a semi-transparent, dark haze. This illusion, Anderson and Winawer propose, is caused by photometric and geometric relationships in the figure that modulate the (non-veridical) perception of *transparency* inside the disk regions: whether a given disk in the demonstration looks white or black depends on the way the visual system uses these relationships for purposes of decomposing the contrasting luminances that define the texture inside the disk into surfaces or layers at different distances in (virtual) depth. In consequence, it is not possible to experience the lightness illusion, while simultaneously experiencing the disks as textured patches on a 2D surface, that is, as *co-planar* with their surround.

Similar remarks can be made with respect to the apparent colours of objects in pictorial space across different virtual illumination conditions. Consider a variant of Dales Purves and Beau Lotto's Rubik's cube colour illusion, in which one cube appears to be lit by a yellow light source and the

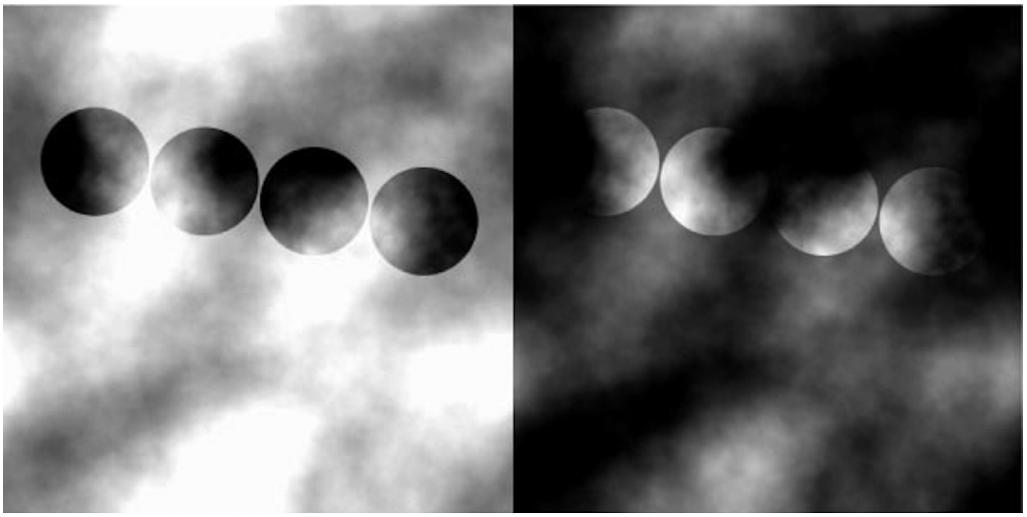

Figure 3.3 Lightness illusion

Reproduced with permission from Anderson and Winawer (2005).

other by a blue light source (Purves *et al*. 2002: 241).[8] What is surprising, when we look at the image, is that although some tiles on the top side of the first cube look blue and some tiles on the top side of the second cube look yellow, the corresponding regions on the pictorial surface are physically identical. Indeed, they are precisely the same shade of grey. This is perceptually evident when a mask is superimposed upon the demonstration, covering the other tiles. The key observation, in the present context, is not just that context affects colour constancy in both physical and pictorial space (Azzouni 2013: 91), but that experiencing the colours of the tiles in phenomenally 3D pictorial space as chromatically different (yellow vs blue) excludes seeing the regions on the 2D pictorial surface in which the tiles are displayed as chromatically identical (grey). It is not possible to enjoy both experiences at the same time. The reader is invited to give it a try![9]

Critics of the dual content theory have sometimes charged its adherents with over-analogizing everyday visual experience to looking at paintings, photographs, and other flat media (Gibson 1979; Smith 2000; Schwitzgebel 2006; Briscoe 2008; Snowdon 2015). Nevertheless, the best case for the dual content theory, it might be thought, is the experience of looking at a picture. For one thing, a picture really *is* a flat patchwork of colours that, when present to sight, elicits the experience as of depth and 3D structure. In addition, it is not clear that criticisms of the dual content theory of everyday visual experience canvassed above straightforwardly apply to a dual content theory of pictorial experience. Indeed, there are two disanalogies between the experience of looking at a voluminous object 'in the flesh' and the experience of looking at an image of such an object that arguably make a dual content theory of pictorial experience more promising.

The first disanalogy concerns the different kinds of optical information available to the visual system when viewing pictures and real-world scenes, respectively. In most cases, sources of information concerning the layout of a real-world scene are rich and consistent enough to support the selection of a single interpretation from the hypothesis space for vision. This explains why it is not possible, holding sources of optical information fixed, to switch from a *disk-slanted-in-depth* interpretation of the retinal image to an *ellipse-in-the-frontal-plane* interpretation in the above example. (In contemporary Bayesian parlance, although both interpretations or hypotheses may have roughly the same prior probability, binocular disparity, texture gradients, vergence cues, and other sources of sensory evidence make one interpretation much more *likely* than the other, and this means that it will be assigned a proportionately higher posterior probability.)

When you look at a picture, by contrast, two conflicting sets of depth cues are typically present in the light reflected (or emitted) to your eye. The first set of cues, as already mentioned, includes sources of optical information that enable your visual system accurately to recover the properties of the opaque, 2D pictorial surface. When you look at *The Peasant Wedding* (Figure 3.1), for example, there is 'surface event' information that specifies

the colours and elliptical shapes of the patches of paint that Bruegel used to depict some dishes of food in the bottom-right quadrant of the canvas. The second set of cues, by contrast, includes sources of monocular spatial information that support the experience as of depth and 3D structure in pictorial space. For example, when viewing Bruegel's painting, there is information that specifies the voluminous shapes, relative sizes, and orientations of the virtual objects that are intended to resemble dishes. Since there is ample information for both of these 3D-scene-interpretations when you look at the painting, it could be argued that both interpretations may be simultaneously reflected in the content of your experience.

There is a second point of disanalogy. One objection to the dual content theory, as we saw, had to do with the fact that it allows for simultaneous attribution of incompatible determinates of the same determinable to the same object. A disk viewed at an oblique angle is supposed to look both *circular* and *elliptical* as well as both *tilted* and *upright* at the same time. This is what Charles Siewert calls the 'Problem of Contradictory Visual Appearances' (Siewert 2006: 5). Matters are rather different, it could be argued, when looking at a picture. Here, one interpretation of the retinal image (I_1) attributes certain properties to the 2D pictorial surface, while the other interpretation (I_2) attributes certain properties to a virtual object in phenomenally 3D pictorial space. In terms of feature-binding theory (Treisman 1996), different features are bound to the surface in I_1 than are bound to the object of pictorial experience in I_2. So there is no contradiction internal to the content of an experience that simultaneously reflects both I_1 and I_2.

How might a Gombrichian respond? With respect to the first putative disanalogy, it is important to observe that there are contexts in which the sources of optical information in the light received from a real-world scene are highly ambiguous, even when the constraining assumptions of the 'beholder's share' are brought to bear on its uptake. In Bayes-speak, there are real-world contexts in which different visual hypotheses have about the same posterior probability. For example, just as it is possible to experience a reversal in orientation when looking at a drawing of the Necker cube (Figure 3.2), it is also possible to experience a reversal in 3D orientation when looking at a cubical wireframe face-to-face (especially when one eye is shut). Another example is the phenomenon of binocular rivalry. In binocular rivalry, the image presented to the left eye is different than the image presented to the right eye, for example, an image of a face and an image of a house (Blake and Logothetis 2002; Alais and Blake 2015). In most cases, however, only one of these objects is visible at a time: the other is suppressed from visual awareness.

The important point is that when sources of optical information in real-world contexts are highly ambiguous, we do not experience different 3D-scene interpretations at the same time. Instead, we *alternate* between competing interpretations. In relevant cases, the visual system does not select a conjoint or 'blended' hypothesis concerning the distal causes of proximal sensory stimulation. We do not experience the cubical wireframe as having

two different orientations at the same time, nor do we experience a blending of the face and house. This makes good sense, from a Bayesian perspective, since conjoint hypotheses such as these have extremely low prior probabilities (Hohwy *et al*. 2008).

The Gombrichian, in short, can maintain that what holds when the information in the light received from a real-world scene is equivocal—that is, when it does not support the selection of a single 3D-scene interpretation—also holds when the information in the light from a patterned, pictorial surface is equivocal. The visual system, in both cases, switches between competing interpretations that 'fit' the available sensory evidence.[10] It does not generate an experience that reflects both interpretations at the same time.

With respect to the second putative disanalogy, the Gombrichian can concede that the two interpretations attribute properties to different objects, viz., the opaque, 2D pictorial surface (I_1), on the one hand, and an object in phenomenally 3D pictorial space (I_2), on the other. This doesn't mean, however, that the two interpretations are consistent. On the contrary, there is a very straightforward reason to think that they are mutually exclusive: I_1 represents the presence an opaque, 2D surface *S* in front of the perceiver, while I_2 represents (non-veridically) an array of objects receding in depth behind *S*. Human perceivers, however, do not have the capacity see through an opaque surface, for example, a sheet of canvas covered with paint, to the world on the other side. If a given solid angle in the visual field appears to contain a non-transparent, 2D surface located at some distance in depth *D*, then it cannot, at the same time, also appear to contain an array of 3D objects at locations more distant in depth than *D*.

3. The Seeing-In Theory

Richard Wollheim is a prominent critic of the Gombrichian account of pictorial experience (1963, 1987). According to Wollheim, when we look at a picture, we not only experience the organization of a virtual 3D scene; we also experience the design present on the 2D pictorial surface. In this sense, pictorial experience—or 'seeing-in', as Wollheim calls it—has two 'aspects' or two 'folds' of representational content:

> Seeing-in is a natural capacity we have—it precedes pictures, though pictures foster it—which allows us, when confronted by certain differentiated surfaces, to have experiences that possess a dual aspect, or 'twofoldness', so that, on the one hand, we are aware of the differentiation of the surface, and, on the other hand, we observe something in front of, or behind, something else.
>
> (1993: 188)

Wollheim argues that while the two 'folds' of content in pictorial experience are clearly different, they are not, contrary to Gombrich, incompatible

(Wollheim 1963: 29). As paradoxical as it may sound, the same solid angle in the visual field, when we look at a picture, is represented twice over: once as encompassing an opaque, two-dimensionally organized surface and once as encompassing a (virtual) three-dimensionally organized scene.

Two different two ways of dispelling the appearance of paradox at the heart of the seeing-in theory have been proposed. The first appeals to the idea that pictorial space and physical space are experienced as entirely unrelated. Wollheim writes that 'there are two distinct dimensions here along which "on," "level with," and "behind" are values: a physical dimension and what we might call a pictorial dimension' (1974: 27). Hence, it would be a phenomenological mistake to describe pictorial experience as representing an opaque, 2D surface *in front* of a 3D scene. This description conflates apparent physical and pictorial depth relations.

The problem with this suggestion is that pictorial space and physical space, while experienced as distinct, are not experienced as entirely unrelated (Hopkins 1998: 195–196; Kulvicki 2009; Lopes 2010). For one thing, objects in pictorial space not only appear to be positioned in certain directions relative to the pictorial point of view, they also appear to be positioned in certain directions relative to the viewing subject's location in physical space. When you watch a film, Lopes notes, 'there is a rich and systematic overlap in the two viewpoint-determining contents—that is, between the represented directions from the camera to depicted objects and directions the objects appear to lie in from the picture viewer in normal viewing conditions' (Lopes 2010: 76). In other words, we can be aware of the direction of an object in pictorial space in both picture- and observer-based frames of reference.

A second, related consideration has to do with the observation that objects in pictorial space sometimes seem to 'follow' the viewer (Gombrich 1972; Goldstein 1979; Koenderink *et al*. 2004; Newall 2015). When the viewer moves in relation to the famous British Army recruiting poster depicting Lord Kitchener, for example, the virtual object she experiences in pictorial space curiously appears to rotate toward her (Gombrich 1961/2000: 113). Koenderink *et al*. write:

> Does this mean that the pictorial object 'rotates along with the observer' as the observer assumes a series of oblique viewing positions by walking along the painting on the wall? . . . *Yes* in terms of the physical space containing the scene, picture, and observer: the pictorial object always squarely faces the observer, it thus looks or points into the observer's visual direction. As the observer changes the visual direction with respect to the picture plane, the looking or pointing direction of the pictorial object in physical and visual space has to rotate with it.
>
> (Koenderink *et al*. 2004: 526)

The main point is that we not only experience the pictorial object's constant orientation in pictorial space vis-à-vis the picture plane, we also experience

its orientation as changing vis-à-vis our own location in physical space. Together, these considerations (but see Kulvicki [2009] for others) suggest that the 'disjoint space' view fails to dispel the appearance of paradox at the heart of Wollheim's account of pictorial experience.

John Kulvicki (2009) enlists our capacity to see through *transparent media* in the service of an alternative account of how twofold pictorial experience is possible. In familiar cases of transparency perception, the visual system distributes light-altering properties, such as colour, transmittance, and glossiness, to surfaces at different distances in depth along the same line of sight. This process, which results in a 'layered' representation of the distal scene, is known as *scission* (Kanizsa 1979; Metelli 1970, 1974; Fleming and Anderson 2004; Anderson and Winawer 2008). When such distribution is unnecessary either because the overlying layer is completely opaque or completely transparent, scission does not occur. As the vision scientist Fabio Metelli writes,

> [i]f all the color goes to the transparent layer, it becomes opaque. If all the color goes to the underlying surface, then the transparent layer becomes invisible. Transparency is perceived only when there is a distribution of the stimulus color to both the [overlying] layer and the [underlying] layer.
>
> (1974: 94)

Kulvicki argues that human capacities for transparency perception allow for the possibility of twofold pictorial experience: the luminance intensities in the retinal image of a picture are used by the visual system to assign properties to the opaque, 2D pictorial surface and, at the same time, to 3D objects that appear to recede in depth behind the surface. It does not matter to this proposal, he says, that

> this kind of layering typically happens when one is confronted with semitransparent objects. . . . What matters is that the visual system has the resources for representing objects and qualities at different distances in one and the same direction, as the literature on transparency perception strongly suggests.
>
> (2009: 393)

The basic problem with this proposal is that the process of scission involves the *distribution* of light-altering properties to surfaces at different distances in depth along the same line of sight. The fundamental question for the visual system, as Fleming and Anderson put it, is, 'How much of the light is due to reflectance of underlying surface, and how much is due to the properties of the overlying layer?' (2004: 1294). However, if, by hypothesis, the pictorial surface is experienced as opaque in twofold pictorial experience (in 'one and the same direction, at a given time, one sees two rather different opaque surfaces, one behind the other: the picture plane and

the receding content' [Kulvicki 2009: 394]), then this just means that no such distribution of light-altering properties can have taken place. Instead of attenuating the paradox at the heart of Wollheim's theory of pictorial experience, the proposal that twofoldness is made possible by our capacity to perceive through transparent media only aggravates it.

Proponents of the seeing-in theory have put forward several reasons to think that twofoldness, however paradoxical it may sound, is a necessary feature of pictorial experience. One reason, Wollheimians maintain, is that twofoldness is required for the aesthetic appreciation of pictures (Wollheim 1980; Nanay 2010). If true, however, this would only entail that we experience certain art pictures in a twofold way, not that twofoldness is necessary for pictorial experience in general (Lopes 1996: 48). More generally, it could also be argued that twofoldness isn't even necessary for purposes of aesthetic evaluation. Why isn't rapidly shifting attention between properties of the pattern on the pictorial surface, for example, the texture of the brushstrokes on a painting, and properties of the virtual object in pictorial space sufficient?

A second reason has to do with evidence that when a picture is viewed from an oblique angle, we do not experience objects in pictorial space as significantly distorted in shape. In other words, shape constancy in pictorial space obtains across changes in perspective. Wollheim argues the best explanation of such constancy is that the viewer not only is aware of the object in pictorial space but also compensates for the orientation of the pictorial surface. Such compensation, he suggests, involves awareness of the 'surface qualities of the representation' (1980: 215–216).

There are three lines of response. First, as Lopes points out, even were it true that surface awareness is necessary for pictorial shape constancy, this wouldn't entail that it is necessary for pictorial experience in general (1996: 49). Second, the compensation process could take place an entirely subpersonal level, using nonconscious information about the orientation of the pictorial surface: Wollheim provides no reason to suppose that compensation depends on conscious perception of surface orientation. Third, there are alternative explanations of pictorial shape constancy in vision science. 'A quite different (and also quite common) view,' as the perceptual psychologist Jan Koenderink and co-authors write,

> is that observers simply don't care, that is to say, disregard (but not in any active sense) the distortions of the retinal image, since the monocular cues as to the structure of the pictorial space are rich enough anyway. In that case, a subsidiary awareness of the picture surface is irrelevant.
>
> (Koenderink *et al*. 2004: 515)

Summarizing the evidence, they conclude that 'there appears to be some (weak) consensus that no "correction" is applied to pictorial space due to obliquely viewed pictures' (Koenderink *et al*. 2004: 526).

According to the seeing-in theory, it is possible to experience the same solid angle in the visual field both as completely filled by an opaque, 2D surface at a single distance in depth *and* as containing an array of 3D objects at different distances in depth. I have shown that two of the main attempts to render this view non-paradoxical confront serious objections, and this speaks in favour of the analogy Gombrich draws between pictorial experience and the experience of looking at an ambiguous figure like the duck–rabbit. In both cases, we 'can train ourselves to switch more rapidly, indeed to oscillate between readings, but we cannot hold conflicting interpretations' (1961/2000: 236).

Before proceeding, however, I would like to make two additional points. First, Gombrich suggests that it is typically possible to alternate at will between awareness of the configuration of colours, textures, and marks on a 2D pictorial surface and awareness of the properties and layout of objects in phenomenally 3D pictorial space, much as it is possible to alternate at will between the different 'aspects' of the Necker cube (Gombrich 1961/2000: 236, 280). There are good reasons, however, to question the generality of this assumption. Whether such voluntary switching is possible when looking at a picture depends, among other things, on the range of depth cues in the light reflected from its surface to the eye. It is quite hard, in fact, to 'unsee' 3D structure in line-drawings that contain appropriate T-junctions, linear perspective, and other depth cues under ordinary, binocular viewing condition (Kennedy 1974; Zeimbekis 2015).[11] It is even harder when standard, surface-specifying cues are eliminated. Looking at a painting through a narrow, monocular aperture often results in a robust impression of depth and 'stereopsis', that is, solid form and immersive space (Kubovy 1986: ch. 12; see the next section). A similar effect is obtained when looking at a single picture through a zograscope (Koenderink *et al.* 2013).[12] These devices eliminate cues that specify the presence and properties of the 2D pictorial surface and, so, render a Gombrichian switch effectively impossible.

Viewing distance also matters. As suggested by the epigraph from *Martin Eden*, it may not be possible to enjoy an experience as of virtual depth and 3D structure when viewing certain paintings at close range.[13] By contrast, it is exceedingly difficult to elude that experience when watching a film from the back row of a movie theater.

These considerations suggest that whether or not it is possible to perform a Gombrichian switch depends on both the sources of information in the light reflected from a pictorial surface as well as the conditions under which the surface is viewed. If this is right, however, then the real problem with Gombrich's deployment of the duck-rabbit analogy, contrary to Wollheim, isn't that it underestimates, but rather that it *overestimates* our ability to see the superficial pattern on a picture's surface as such.

The second point is that when a Gombrichian switch between competing interpretations of the retinal image of a picture *is* possible, iconic and/or visual short-term memory of 'pre-switch' interpretations may support the

illusion of simultaneity. That is, instead of simply oscillating between an experience as of a pattern of brushstrokes on a 2D canvas (I_1) and an experience as of an object or scene in phenomenally 3D pictorial space (I_2), as Gombrich suggests, the viewer may oscillate between I_1 + a sensory, visual memory of I_2 and I_2 + a sensory, visual memory of I_1. In consequence, she may deceptively seem to enjoy both visual experiences (I_1 and I_2) at once. Also at work may be a kind of "immanence" (Minsky 1986) or "refrigerator light" illusion (Block 2001), in which the viewer mistakes the potential presence in visual consciousness of a certain set of properties for their actual presence. As an example of the refrigerator light illusion, here is Eric Schwitzgebel's (2008) explanation of why naïve introspectors tend to overestimate their visual acuity outside the central, foveal region of their visual field:

> Here's the root of the mistake, I suspect: When the thought occurs to you to reflect on some part of your visual phenomenology, you normally move your eyes (or "foveate") in that direction. Consequently, wherever you think to attend, within a certain range of natural foveal movement, you find the clarity and precision of foveal vision. It's as though you look at your desk and ask yourself: Is the stapler clear? Yes. The pen? Yes. The artificial wood grain between them and the mouse pad? Yes—each time looking directly at the object in question—and then you conclude that they're all clear simultaneously.
>
> (Schwitzgebel 2008: 255; also see Dennett 1969: 139–140)

Similarly, I would suggest, our ability to switch attention without significant delay between the colours, textures, and other superficial properties visible within some pictorial surface region *R*, on the one hand, and the properties of the virtual 3D object or scene displayed by *R*, on the other, may contribute to the illusion that both sets of properties are present in visual consciousness at the same time.

Consider in this connection the view that certain pictures elicit a twofold experience that Dominic Lopes refers to as 'design seeing' (2005: 28). When we engage in design seeing, according to Lopes, we see the configuration of design features visible on a picture's 2D surface as 'undergirding' or as 'responsible for' the very experience of depth and 3D structure that those features elicit in us. Pictures that elicit design seeing 'wear the process of depiction on their sleeves' (Lopes 2005: 52). This view is controversial, however, not only because it takes the possibility of twofoldness for granted but also because it presupposes that the contents of visual experience are rich enough to represent pictorial design features as such. That is, it presupposes that visual experiences (and not only visually based beliefs) have capacity to represent the high-level property *pictorial design feature*.

I do not want to take a stand here on the dispute between 'rich' and 'thin' theories of visual representational content (but see Prinz 2006; Siegel

2010, 2016; Briscoe 2015; and Byrne 2016). For present purposes, the point is only that the visual-memory/attentional-switching proposal provides resources to explain why looking at a picture may sometimes deceptively appear to involve design seeing. Because, in relevant cases, we have the capacity to switch attention between the set of design features visible within some pictorial surface region *R* and properties of the virtual, 3D object or scene displayed by *R* and, in addition, because our experience of the design features visible within *R* may be, so to speak, coloured by a concurrent, visual memory of that virtual object or scene, it may seem to us as though our experience of those design features represented them *as such*—and not merely as a superficial configuration of marks, colours, textures, and other low-level properties.

4. Toward a Weakly Onefold Theory of Pictorial Experience

Dominic Lopes (1996) refers to the view that twofoldness is essential to pictorial experience as 'strong twofoldness' and, correspondingly, to the view that twofoldness is merely consistent with pictorial experience as 'weak twofoldness'. Gombrich, by contrast, is sometimes interpreted as defending what might be called a *strongly onefold* conception of pictorial experience. According to strong onefoldness, it is not psychologically possible to experience *any* properties of the 2D pictorial surface while simultaneously experiencing properties of a virtual scene in phenomenally 3D pictorial space. To the extent that we are aware of virtual depth and 3D structure when viewing *The Peasant Wedding* (Figure 3.1), for example, we are effectively blind to the superficial pattern on the painting's canvas.

Strong onefoldness is a core component of the so-called illusion theory of pictorial experience. The illusion theory goes far beyond the claim that pictorial experience and seeing face-to-face are experiences of the same psychological kind (the Continuity Hypothesis). It further maintains that pictures elicit experiences that cannot be introspectively distinguished from experiences of seeing the objects that they depict face-to-face. According to the illusion theory, Lopes writes, 'one sees *O* in a picture when and only when one's experience as of *O* when looking at the picture is phenomenally indistinguishable from a face-to-face experience of *O*' (2005: 30). But, if the experience of looking at a picture of an object is supposed to be phenomenologically on all fours with the experience of seeing an object with the same properties face-to-face, then the former experience evidently cannot involve seeing the properties of the picture's surface. Not surprisingly, the paradigm case of pictorial experience for the illusion theory is the experience of seeing a *trompe l'oeil* painting from the appropriate station point.

The possibility of pictures that 'fool the eye' indicates that pictorial experience is, at least sometimes, strongly onefold. Putting *trompe l'oeil* paintings to the side, however, most pictures do not elicit experiences that are introspectively difficult to distinguish from experiences of actually seeing

their *depicta*. As Gombrich points out, 'we rarely get into situations in which the eye is actually deceived' (1961/2000: 246). One reason is straightforward. When we look at a picture of some high-level kind of *F*, for example, a woman, or a tree, or a clock, the properties attributed by our visual system to the object we experience are very often different from those that would be attributed to an actual *F*, when seen from the relevant point of view. Indeed, the intentional object of our experience may appear to have properties that no actual *F* could have (think of Magritte's surrealist paintings) and/or to lack properties that no actual *F* could lack (think of Picasso's highly abstract line drawings).

These considerations present a serious challenge to the illusion theory, but they do not threaten the Continuity Hypothesis. That the content of the experience caused by a picture of a horse typically does not match the content of an experience that might have been caused by an actual horse does not conflict with the claim that they are experiences of the same psychological kind. In this respect, after all, the experience elicited by a picture of a horse is completely on par with the experience of seeing a 3D sculpture or model of a horse in the flesh.

I have indicated one reason why pictorial experience is typically non-deceptive: the properties attributed by pictorial experience to its intentional object are in many cases different from those that the depicted object would be seen to have when confronted face-to-face. There is a more profound respect, however, in which the experience elicited by a picture is normally distinguishable from the experience of actually seeing the object that it depicts. Robert Hopkins writes,

> In some way, when I see a woman in a painting, I am visually aware of a woman. . . . I am presented with a woman, but not so as to suggest that *that* is what is really there. Unlike perceptual consciousness, this awareness is non-committal about the reality of its objects. In this respect, if no other, pictorial consciousness is like visualizing.
>
> (Hopkins 2012b: 434)

Similarly, Michael Martin observes that when we look at pictures, we experience objects that manifestly lack 'solidity' and a 'self-standing appearance'. The objects of pictorial experience, in contrast with objects seen face-to-face, are experienced as '*mere* visibilia' (Martin 2012: 342). Other authors in this connection refer to a feeling of 'presence' or 'reality' that accompanies seeing face-to-face but is conspicuously absent when looking at a picture (Sartre 1940/2004; Michotte 1960; Matthen 2005; Wiesing 2010).

That the experience elicited by a picture of an object and the experience of seeing the depicted object face-to-face do not typically match in content, as we saw, does not threaten the Continuity Hypothesis. That pictorial experience typically does not dispose us to believe that its intentional object is real—'something out there'—by contrast, presents a serious challenge. As

Hopkins points out, if the claim that pictorial experience is an experience of the same psychological kind as the experience of seeing face-to-face is to be informative, then the respects in which former is phenomenally distinguishable from the latter must be limited to differences in the properties—the shapes, sizes, colours, and so forth—that they respectively attribute to their intentional objects:

> The move from presenting [the intentional object] *O* as real to no longer doing so hardly fits that bill. One central difference between seeming to see something and visualizing it is that the former necessarily presents its object as real. Failure to do that secures that [pictorial experience] cannot be [a] visual experience as of *O*, and opens up the possibility that it is visualizing.
>
> (Hopkins 2012a: 654–655)

These considerations not only put pressure on the illusion theory, they also put pressure on the Continuity Hypothesis. If everyday visual experience typically disposes its subject to believe that its intentional objects are real, but pictorial experience does not, then this is seemingly a good reason to suppose that we are dealing with experiences that not only contrast in representational content but also in psychological kind.

The remainder of this section has two aims. The first is to motivate an account of the structure of everyday, non-deceptive pictorial experience that I refer to as *weak onefoldness*. Everyday pictorial experience, according to the account, is onefold in the sense that its content reflects a single, consistent 3D-scene interpretation of the retinal image. It does not represent the same solid angle in the visual field twice over—once as filled by an opaque, two-dimensionally organized surface and once as filled by a three-dimensionally organized scene. Everyday pictorial experience, however, typically is only *weakly* onefold in the sense that it attributes properties to the pictorial surface and to objects in pictorial space at the same time. It represents a single scene with both real and virtual constituents—a scene that straddles the boundary between physical and pictorial space. On this approach, surface representation in pictorial experience is not an all or nothing affair. Having the experience as of virtual depth and 3D structure, when looking at a picture, I argue, excludes representing some, but not *all* of its superficial properties. The important consequence, in the present context, is that is possible to embrace Gombrich's requirement of a single, consistent, 3D-scene interpretation while rejecting strong onefoldness and the illusion theory.

The second aim is to reconcile the claim that pictorial experience and seeing face-to-face are psychologically continuous with the observation that the (virtual) objects of pictorial experience normally do not appear to be present in the physical environment. An empirically motivated account of *stereopsis* developed by Dhanraj Vishwanath (2014), I argue, makes such reconciliation possible.

In order to motivate weak onefoldness, it is necessary to demonstrate that although pictorial experience is structured by just one layer or 'fold' of representational content, that content can consistently attribute certain combinations of properties to the 2D pictorial surface and to objects in phenomenally 3D pictorial space at the same time. There are at least two different ways, I propose, in which this is possible.

First, while it may not be possible simultaneously to experience the same solid angle in the visual field as filled by an array of voluminous 3D objects that recede in depth *and* by a single, non-transparent 2D surface, it does seem possible, when looking at a picture, to divide attention between the region of pictorial space contained within some solid visual angle θ and the distribution of pictorial surface properties contained within some different solid visual angle φ. For example, we can divide our attention between the 3D shape of an object displayed in one region of a canvas and the facture or colour of the paint laid down in some other. If so, then we may be visually aware of certain pictorial-space properties and certain pictorial-surface properties at the same time. None of the considerations adduced earlier in support of Gombrich's account of pictorial experience rule out distributing our attention and the content of our experience across the pictorial space/physical space frontier in this manner.

Second, it is possible to experience regions on a pictorial surface as opaque and as partially occluding objects in pictorial space. Words printed on a pictorial surface, for example, may be visually experienced as occluding figures on a more distant, three-dimensionally organized background, and when this is the case, we experience properties of the real-world surface and virtual scene at the same time. Posters, movies with subtitles, advertisements, and other pictures that incorporate textual elements elicit experiences that plausibly fit this description. They provide an existence proof of the possibility of weakly onefold pictorial experience.

As an example, consider the poster for the Buster Keaton film *Sherlock, Jr.* reproduced in Figure 3.4. Some edges, including those belonging to the letters in the title of the film are represented by the visual system as surface events and, so, are assigned to locations on the pictorial surface. Other edges, in contrast, are assigned to locations beyond the surface in pictorial space. The contour of the virtual object resembling Keaton's head, for example, appears to occlude the lowermost edges of some of the letters in the name 'Keaton', with the result that the entire name is dragged back in depth. It appears to float inside the virtual, 3D scene displayed by the poster, while the title *Sherlock, Jr.* rests on the plane of the poster's surface. Contradiction is avoided because no edge is interpreted in both ways at once. The overall experience elicited by the poster, I would suggest, is one in which certain textual elements appear partially to occlude objects that recede in depth from the plane of the pictorial surface.

There are constraints, of course, on which experienced combinations of 2D surface properties and virtual, 3D scene properties are psychologically

Figure 3.4 Poster for *Sherlock, Jr.* (1924)

possible. Consider once again a drawing of the Necker cube (Figure 3.2). There are eight vertices or 'y-junctions' present in the retinal image of the drawing, each of which can be interpreted in at least three different ways: as convex, that is, pointing toward the viewer; as concave, that is, pointing away from the viewer; or as co-planar, that is, as the intersection of three line segments on the same plane of depth as the pictorial surface. Each local interpretation constrains how the visual system interprets the remaining seven y-junctions in the image. Interpreting y-junction *a* as convex, for instance, constrains the visual system to interpret *b*, *c*, *d* as pointing in the same direction with the result that face *abcd* is experienced as in front of face *efgh*. Interpreting *a* as concave results in the reverse depth ordering. Interpreting *a* as co-planar with the pictorial surface results in a globally 'flat' interpretation, in which neither *abcd* nor *efgh* is experienced as in front

of the other. Each local interpretation, in short, constrains how the visual system interprets the other edges present in the retinal image (Albert and Hoffman 2000; Cooper 2008). For present purposes, the main point is that the visual system processes 3D structure in a holistic way. In the pictorial case, this means that whether a given edge in the retinal image is interpreted as an event on the 2D pictorial surface or as the contour of a (virtual) 3D object typically has consequences for how other nearby edges in the retinal image are interpreted.

I have argued the visual system can consistently attribute certain properties to the pictorial surface and to objects that appear beyond the surface in pictorial space at the same time. When it does, the experience elicited by a picture is weakly onefold. It represents a single, integrated scene containing both real and virtual elements. If this is right, however, then it possible to embrace the Continuity Hypothesis and Gombrich's requirement of a single, consistent, 3D-scene-interpretation while rejecting strong onefoldness.

I turn, now, to the observation that pictorial experience does not typically dispose the viewer to believe that its intentional object is real, that is, present to sight in physical space. Does this observation, as Hopkins suggests, speak against the Continuity Hypothesis?

It is helpful to begin by reflecting on some of the very different reasons why visualizing and pictorial experience do not tempt the subject to believe that she is perceptually related to an object in the physical environment. The primary reason that visualizing is 'non-committal' about the real presence of its objects is that its content is subject to the will (Sartre 1940/2004; Wittgenstein 1953; McGinn 2004).[14] Visualizing, like thinking, is something we do, not something that happens to us. Accordingly, we do not adopt an attitude of *observation* toward its intentional objects. This is not to deny that we can sometimes learn about the world by engaging in visualization: a subject, for example, could imagine what her flat would look like if all the walls in it were transparent and thereby come to realize for the first time that certain windows were opposite each other.[15] The right point is that the contents of visualization are either imported from the subject's memory and belief systems (as when she imagines the layout of the flat) or derived from certain manipulations of that information (as when she counterfactually imagines that the walls in the flat are transparent). It is in this sense that the objects of visualization are 'not in a position to feed new information into our cognitive system' (McGinn 2004: 20).

The contents of pictorial experience in contrast with the contents of visualization are not in general subject to the will. And, like the experience of seeing face-to-face, the content of pictorial experience is systematically guided by the flow of information from the external environment. A pictorial surface will elicit an experience as of an object with a specific 3D shape Σ, for instance, only if it reflects light to the eye of the same type as would elicit experience as of a Σ-shaped object when reflected from a real-world scene.[16] '[P]erceiving depth in pictures and perceiving depth in

the real world', as the psychologist James Cutting puts it, 'are cut from the same informational cloth' (2003: 236). Pictures, in short, can elicit experiences containing genuinely novel, visual content. They can surprise us and invite what McGinn refers to as an attitude of cognitive openness. In this respect, experiencing an object with certain properties in pictorial space is fundamentally unlike visualizing an object with the same properties. Nonetheless, pictorial experience doesn't so much as tempt the perceiver to judge that its object is really there in front of her. Why is this the case?

In order to answer this question, it is necessary to begin by introducing two distinctions. First, following Vishwanath (2014), we need to distinguish between having the experience as of depth and 3D structure, on the one hand, and having the experience of *stereopsis* (from the Greek for *solid* and *appearance*), on the other. Both pictures and real-world scenes elicit the former experience. Vishwanath characterizes the experience of stereopsis, in contrast, as the 'vivid impression of tangible solid form, immersive negative space and realness that obtains under certain viewing and stimulus conditions' (Vishwanath 2014: 153). Pictures, as already observed, typically elicit the former experience without eliciting the latter. When we look at *The Peasant Wedding*, for example, we experience the layout of a complexly organized, 3D scene, but we are hardly disposed to judge that the objects we experience are really there in front us.[17]

Second, it is also necessary to distinguish between the sources of *relative depth* information and sources of *absolute distance* information available to the visual system (Landy *et al*. 1995; Banks *et al*. 2011). Independently variable sources of relative depth information include, but are not limited to, occlusion, texture gradients, shading, the kinetic depth effect, height in the visual field, and binocular disparity. These relative cues specify, among other things, depth ratios between objects in a visible scene, for example, that object *a* is twice as far away as object *b*, and simple ordinal relations, for example, that *a* is behind *b* but in front of *c*. They do not, however, specify the absolute distance of an object from the eye in either conventional units such as meters or non-conventional units such as eye heights (Sedgwick 1986; Bennett 2011). Sources of absolute or 'metric' distance information comprise motion parallax, defocus blur, and the oculomotor cues of convergence and accommodation. In order to integrate these two different types of spatial information for purposes of representing the 3D layout of a visible scene, relative cues must be scaled or 'promoted' by absolute cues so that the depth values they provide are commensurate (Landy *et al*. 1995: 392). For example, the convergence angle of the eyes is a source of absolute information about fixation distance, which, when combined with an estimate of the intraocular separation, can be used to promote binocular disparity to an absolute distance cue.[18]

According to what Vishwanath calls the *absolute depth scaling hypothesis* (ADSH), the experience of stereopsis obtains under real-world, binocular viewing conditions because both sources of relative depth and absolute distance

information typically are available to the visual system. Only relative depth information, however, is typically available for objects in pictorial space, and this explains why pictorial experience is not accompanied by the impression of stereopsis:

> When a picture is viewed binocularly, distance cues such as binocular convergence, vertical disparity and the accommodative state of the lens specify the distance of the visible picture surface, so there are no optical distance cues that specify the distance of pictorial objects. Pictorial depth cues, such as shading, perspective, and interposition, can specify the 3-D shape and relative layout of objects in the depicted scene, but without distance information, these cues cannot be scaled to derive absolute depth or size.
>
> (Vishwanath 2014: 158–159)

A number of philosophers have observed that pictorial experience does not represent its objects as standing in fully determinate, egocentric spatial relations to the observer herself (Carroll 1996; Cohen and Meskin 2004; Matthen 2005; Nanay 2014). In consequence, although the observer may be aware of depth and detailed 3D structure when looking at a picture, the virtual scene it displays does not ordinarily appear to be a potential arena for performing visuomotor actions. The ADSH provides a psychophysical explanation of *why* pictures are, as Cohen and Meskin (2004) put it, 'spatially agnostic informants': sources of absolute distance information in the light received from a pictorial surface are assigned to the 2D pictorial surface, rather than to objects located in phenomenally 3D pictorial space.[19]

One source of evidence for the ADSH is its ability to explain why the experience of stereopsis is successfully induced when paintings, photographs, and other types of pictures are viewed with one eye through a peephole (for others, see Vishwanath 2014). Under these conditions, neither vergence nor disparity is available as a source of information for the absolute distance of the pictorial surface. Accommodation-based distance information, however, remains, and, in the absence of a visible pictorial surface, Vishwanath proposes, it is reassigned to objects in pictorial space, enabling the visual system to compute estimates of their absolute distance in depth and intrinsic size. The ADSH predicts that the reassignment of accommodation-based distance information to pictorial objects should make them appear to be relatively close to the observer—that is, as located at roughly the actual distance of the pictorial surface from the eye—and, hence, as smaller than the real-world objects they are intended to resemble. (Indeed, accommodation is an effective range-finder only for objects in nearby space [Cutting and Vishton 1995].) Subjective reports concerning the effects of monocular aperture viewing bear this prediction out (Vishwanath and Hibbard 2013). The cars and buildings in a photograph of a cityscape, when viewed through

a narrow peephole, often have the appearance of being miniaturized and toy-like. This experience is sometimes referred to as the 'diorama effect'.[20]

As Hopkins argues, the claim that pictorial experience and seeing face-to-face are experiences of the same psychological kind is informative only if the respects in which the two experiences are phenomenally distinguishable are at the level of representational content: they must be limited to differences in attributed visual appearance properties. On Vishwanath's account, however, the reason that pictorial experience does not dispose the observer to believe she is seeing a real object is that it fails to specify the object's location at some absolutely scaled distance in depth. But, if this right, then the relevant contrast with seeing face-to-face is best understood as a difference at the level of representational content rather than a difference at the level of psychological kind. Just as visual experiences can be more or less determinate with respect to properties such as shape, size, and colour (Tye 2003)—an object present in peripheral vision, for example, may be experienced as dark in colour or as elongated in shape but not as instantiating a specific dark colour or a specific elongated shape—they can also be more or less determinate with respect to the distance of an object *vis-à-vis* the observer herself. In the limit, conscious vision simply may not offer comment on an object's absolute distance in depth. In short, if the ADSH is correct, then the fact that the objects of pictorial experience do not appear to be present in the same space as the surface in which they are displayed does not speak against the psychological continuity of pictorial experience with seeing face-to-face.

5. Conclusion

I will conclude this chapter with a speculative proposal. According to Gombrich's exclusivity requirement, the content of visual experience always reflects a single, consistent, 3D-scene interpretation of the retinal image. In the last section, I argued that weakly onefold pictorial experiences satisfy the exclusivity requirement, and I developed this claim with two examples.

Here is a third possible way in which pictorial experience *could* be weakly onefold. The ADSH maintains that, when we look at a picture, absolute distance in depth estimates are only computed for the picture's surface. Conscious vision is silent when it comes to the absolute distance of an object positioned 'beyond' the surface in pictorial space. In this respect, the content of our experience when we look at a picture is significantly less determinate than the content of our experience when we engage in seeing face-to-face. Now, one way in which the experience elicited by a picture could represent the presence of the picture's surface *and* the configuration of objects in pictorial space within the same solid angle in the visual field at the same time would be if it also represented the surface in a highly indeterminate way, in particular, if it represented the viewer-relative absolute distance and orientation of the surface, as well as some of its intrinsic geometric properties (curvature, shape, size), but did not attribute any *colour properties* to

it. Just as the visual system does not offer comment on the absolute distance of an object embedded in pictorial space—this explains why the object never appears closer to us when we move in its direction (Matthen 2005: 316–317)—it may not offer comment on the colour(s) of the surface in which the object is displayed.

This description, I would suggest, phenomenologically fits the experience of looking at certain photographs and naturalistic paintings, especially at a distance, as well as the experience of watching movies. In 'the normal case of watching a movie', Colin McGinn writes, 'we don't focus our attention on [the] fleeting patterns of light—we, as it were, look right *through* them' (McGinn 2005: 17). But this does not mean that visual awareness of the surface drops out entirely. The colour properties we experience when we watch a movie are assigned by our visual system to virtual objects on the 'other side' of the screen, but we are still aware of the screen's presence and typically have no difficulty estimating its distance from us or its orientation.

This proposal is not as counterintuitive as it may sound. In a recent treatment, Fiona Macpherson (2015) critically evaluates the traditional, Aristotelian view that visual experience as of an object necessarily represents the object's colour(s). Drawing on evidence concerning tactile-visual sensory substitution, amodal completion, and type 2 blindsight, she argues that there is no good reason to deny that visual awareness of 'pure distal form without colour' is possible. Contrary to philosophical tradition, colour may not be a structural feature of object-representing visual experience. Certain cases of picture perception, if I am right, put another arrow in Macpherson's quiver.

To be clear, I am not proposing that in relevant cases we experience the pictorial surface as transparent to the virtual 3D scene. Recall Metelli's constraints on transparency perception:

> If all the color goes to the transparent layer, it becomes opaque. If all the color goes to the underlying surface, then the transparent layer becomes invisible. Transparency is perceived only when there is a distribution of the stimulus color to both the [overlying] layer and the [underlying] layer. (1974: 94)

In the experience that I am characterizing, however, there is no *distribution* of colour across layers. Instead, all the colour goes to the underlying layer, that is, to the object (or array of objects) in pictorial space. But this does not entail that the pictorial surface is invisible, like a perfectly clear pane of glass. Instead, we may visually experience the surface in a highly indeterminate way: as present at some absolute distance in depth and as having such-and-such orientation relative to our line of sight but not as having any colour properties of its own. If so, then there would be no inconsistency internal to an experience that represented *both* the surface and (non-veridically) a 3D scene receding in depth behind it.

Michael Newall develops a similar view in a recent treatment. According to Newall, when we look at certain photographs, '*all* the picture surface's

visible properties are attributed to the subject matter and the viewer loses all visual awareness of the picture surface' (2015: 144). His example involves the photographic image of a glass of milk that appears on the cover of some editions of the novel *A Clockwork Orange*:

> The milk is depicted by a white, or slightly grey, colour. Here my claim is that this colour is wholly attributed to the subject matter—the milk—which appears as if a little behind the picture surface. While we have this experience, the picture surface appears as if it lacks all its colour properties. It appears, one might say, transparent to the point of invisibility.
>
> (Newall 2015: 145)

We can, to be sure, also experience the white colour as belonging to the pictorial surface, Newall allows, but this requires a Gombrichian switch.

My proposal in this concluding section owes a lot to Newall's discussion of this example. It differs from his own proposal, however, in that it denies that in relevant cases the pictorial surface is completely invisible. Instead, I claim that we may be aware of the surface but in an uncharacteristically indeterminate way: the surface may be represented as having a certain absolute distance in depth and orientation, as well as certain intrinsic, geometric properties, but not as having any colour properties. This seems phenomenologically more plausible than the claim that the surface is altogether absent from visual awareness since we can attend to the changing distance or orientation of a photograph like the one discussed above, while simultaneously paying attention to the virtual 3D object or scene that it displays.[21]

Notes

1. Two clarificatory remarks. First, there are many different types of pictures and any given picture can elicit many different types of mental states in its viewers. Philosophers of perception and vision scientists, however, are specifically interested in those pictures that elicit an experience as of depth and 3D structure (for a useful, interdisciplinary overview, see the essays collected in Hecht *et al*. 2003). Accordingly, my focus is on pictures that do elicit this distinctive experience. Although I do not argue for the claim here, I think that they constitute a fairly unified artefactual kind. Second, the point of using the 'experience as of' locution is to bracket the question concerning whether the experience in question is veridical. A visual experience as of a Σ-shaped object, for example, may be an experience in which the perceiver actually sees a Σ-shaped object or an experience in which she only *seems* to see a Σ-shaped object.
2. There are ways of developing Walton's imagining seeing account on which it is an elaboration of (A). Looking at Kasimir Malevich's *Suprematist Painting*, for example, may elicit a non-veridical visual experience *as of* a yellow rectangle in front of a green rectangle in front of a black trapezoid, which, in turn, may prompt the viewer to imagine that she is actually seeing a three-dimensionally organized scene with the properties represented in the former non-imaginative experience (Walton 1990: 56). In this sort of case, the content of the experience of imagining seeing is asymmetrically dependent on the content of an experience of the type described in (A).

3. What about Escher lithographs such as *Belvedere*, the *Devil's Tuning Fork*, and other 'impossible figures'? These images display 3D structures with locally consistent parts that are assembled in a globally inconsistent way. It is far from clear, however, that we visually experience the global inconsistency of an impossible figure *as such*. Instead, we may see *that* the building depicted by Escher's *Belvedere* is impossible—in Dretske's (1969) epistemic sense of 'seeing that'—by serially inspecting its locally consistent parts and drawing the conclusion that they cannot be coherently assembled (Gombrich 1978; Bayne 2010: 54–55). Julian Hochberg's work is especially helpful here (see the essays and commentaries collected in Peterson *et al*. [2007]). According to Hochberg, locally consistent regions of an impossible figure registered across multiple fixations are not normally compared to each other directly. Instead, we notice global inconsistencies when the internal model or 'schema' of the virtual object's 3D structure set up by the depth cues available in one region of the picture is found to conflict with information about the object's structure available in some other subsequently fixated region. In keeping with this account, perceivers are much more likely to experience an impossible figure as a flat 2D form when globally inconsistent regions are close enough on the pictorial surface to be compared at a single glance.
4. Hatfield (1990) and Smith (2000) provide helpful historical overviews.
5. Compare Hume's claim that 'all bodies, which discover themselves to the eye appear as if painted on a plain surface' (1740/1978: 56).
6. Although multistable perceptual experiences, as we have seen, can occur in which the selected hypothesis alternates from one moment to the next, depending on the allocation of attention and other factors. A clear example here is the flip in depth assignments when viewing a drawing of the Necker cube (Figure 3.2).
7. The affinities described in this paragraph are no accident: Gombrich cites as major influences on his account of pictorial experience Tolman and Brunswik's 'The Organism and the Causal Texture of the Environment' (1935) and Hayek's *The Sensory Order* (1952), both of which anticipate ideas central to the Bayesian approach.
8. An online demonstration is available here: www.lottolab.org.
9. Might consciously experiencing the chromatic properties instantiated by a picture's surface nonetheless have epistemic or computational priority relative to perceiving the way chromatic properties appear to be distributed in pictorial space? Recent metacontrast-masked priming experiments conducted by Liam Norman and colleagues (2014) provide empirical evidence that speaks against this assessment.
10. This is not to deny, of course, that it is usually possible to falsify one (or more) of these interpretations through movement and active exploration in real-world cases (Gombrich 1961/2000: 274–275), that is, by 'sculpting the flow' of incoming information from the environment (Clark 2015: ch. 4.)
11. 'Everyone knows', Wittgenstein remarks, 'that a cube which is clearly drawn will be seen three-dimensionally. One might not even be able to describe what one sees in anything other than three-dimensional terms' (1980: 85).
12. This is sometimes referred to as paradoxical monocular stereoscopy (Schlosberg 1941; Koenderink *et al*. 1994; Koenderink *et al*. 2013).
13. Compare Kenneth Clark's story about 'stalking' Velázquez's painting *Las Meninas*, discussed in the introduction to *Art and Illusion* (Gombrich 1961/2000: 6).
14. This is not to claim that visual imagining in general is subject to the will. See Briscoe (Forthcoming) for discussion.
15. I am grateful to Fiona Macpherson for this example.
16. Many different token light arrays, it should be emphasized, can convey optical information about the same spatial attribute. The arrays of light respectively reflected from a cubical wire framework seen face-to-face, a photograph of the framework, and a line drawing of the framework, for example, can convey

substantially the same information about 3D shape and orientation despite the numerous photometric differences between them (Gibson 1979; Kennedy 1974).

17. Stereopsis is often taken to require (or even to be identical with) the computation of depth from binocular disparities. Vishwanath points out that this is an error since the experience of stereopsis can be readily induced under monocular viewing conditions: when you close one eye, objects in front of you do not seem to be less real or 'out there'.
18. There is significant debate when it comes to specifying the *units* in terms of which visual experience 'measures' absolute distance and size. Robert Schwartz writes: 'It has long been recognized that the pictorial cues cannot indicate absolute spatial measures. . . . Of the traditional cues only the nonvisual motor cues of convergence and accommodation might seem to vary directly and unambiguously with distance. . . . Still, in order to evaluate absolute distance it is not enough to have a cue *K* that varies directly and unambiguously with distance. In addition, we need a scheme for assigning absolute-distance meaning to the values of *K*. We must know how much distance goes with so much *K*' (Schwartz 2006: 23–24). One empirically well-motivated view is that visual experience represents absolute distance in terms of the height of the perceiver's point of view from the ground plane, that is, in eye-height units. For discussion, see Bennett (2011), Firestone (2013), and Briscoe (2015).
19. Several philosophers including Mohan Matthen (2005: ch. 13) and Bence Nanay (2010, 2014) have proposed that objects in pictorial space are represented by the ventral visual information processing stream in the brain, but are not normally represented by the action-guiding, dorsal processing stream. This, they argue, explains why the 'feeling of presence' does not typically attend pictorial experience. The ADSH should be distinguished from the Matthen-Nanay proposal. That absolute distance information required for programming motor actions is not available for the objects of ordinary, non-deceptive pictorial experience, does not entail that the dorsal stream is completely blind to depth and 3D structure in pictorial space (see Briscoe 2016).
20. As Kubovy (1986) points out, however, depth of field is inversely proportional to peephole size. In consequence, a picture seen through a very narrow aperture 'would be nicely in focus even if the eye accommodated so that its focus plane would be at the distance one might expect the walls of a real room to be' (1986: 36). If this is right, then the reassignment of accommodation-based distance information to objects in pictorial space, under peephole viewing conditions, may place them at locations significantly further away from the perceiver than the actual location of the 2D pictorial surface. The location of a peephole relative to the pictorial surface and to the physical ground plane also matters. In her study of 17th-century Dutch perspective boxes, Susan Koslow writes, 'One of the most surprising effects of a perspective box according to Samuel van Hoogstraten was that a figure only a finger's length appeared as large as life. This illusion was achieved because the figure was viewed from its own eye-level. As a consequence, the spectator felt himself included within the projected space of the perspective box' (Koslow 1967: 38).
21. I am grateful to David Bennett, Derek Brown, Clotilde Calabi, E.J. Green, Fiona Macpherson, Lisa Mosier, Paul Noordhof, Paolo Spinicci, Tom Stoneham, and Alberto Voltolini for helpful discussions of the ideas presented in this chapter.

References

Alais, D. and Blake, R. (2015). Binocular rivalry and perceptual ambiguity. In J. Wagemans (ed.), *The Oxford Handbook of Perceptual Organization*. New York: Oxford University Press.

Albert, M. and Hoffman, D. (2000). The generic-viewpoint assumption and illusory contours. *Perception* 29: 303–312.

Ames, A. (1925). The illusion of depth from single pictures. *Journal of the Optical Society of America* 10: 137–147.

Anderson, B. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception* 26: 419–453.

Anderson, B. and Winawer, J. (2005). Image segmentation and lightness perception. *Nature* 434: 79–83.

——— (2008). Layered image representations and the computation of surface lightness. *Journal of Vision* 8: 1–22.

Azzouni, J. (2013). *Semantic Perception: How the Illusion of a Common Language Arises and Persists*. Oxford: Oxford University Press.

Banks, M. S., Burge, J., and Held, R. T. (2011). The statistical relationship between depth, visual cues, and human perception. In J. Trommershäuser, K. Kording, and M. Landy (eds.), *Sensory Cue Integration*. Oxford: Oxford University Press, 195–223.

Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press.

Bennett, D. J. (2011). How the world is measured up in size experience. *Philosophy and Phenomenological Research* 83: 345–365.

Blake, R. and Logothetis, N. (2002). Visual competition. *Nature Review of Neuroscience* 3: 13–21.

Block, N. (2001). Paradox and cross purposes in recent work on consciousness. *Cognition* 79: 197–219.

Briscoe, R. (2008). Vision, action, and make-perceive. *Mind and Language* 23: 457–497.

——— (2015). Cognitive penetration and the reach of phenomenal content. In A. Raftopoulos and J. Zeimbeikis (eds.), *The Cognitive Penetrability of Perception: New Philosophical Perspectives*. Oxford: Oxford University Press, 174–199.

——— (2016). Depiction, pictorial experience, and vision science. In Christopher Hill and Brian McLaughlin (eds.), *Philosophical Topics* 44(2): 41–87.

——— (Forthcoming). Superimposed mental imagery: On the uses of make-perceive. In F. Macpherson and F. Dorsch (eds.), *Perceptual Memory and Perceptual Imagination*. Oxford: Oxford University Press.

Brogaard, B. (2012). Vision for action and the contents of perception. *The Journal of Philosophy* 109: 569–587.

Budd, M. (2008). *Aesthetic Essays*. Oxford: Oxford University Press.

——— (1992/2008). On looking at a picture. In J. Hopkins and A. Savile (eds.), *Psychoanalysis, Mind and Art: Perspectives on Richard Wollheim*. Oxford: Blackwell, 259–280. Reprinted in Budd (2008).

——— (1993/2008). How pictures look. In D. Knowles and J. Skorupski (eds.), *Virtue and Taste: Essays on Politics, Ethics, and Aesthetics*. Oxford: Basil Blackwell, 154–175. Reprinted in Budd (2008).

Byrne, A. (2016). Rich or thin? In B. Nanay (ed.), *Contemporary Debates in the Philosophy of Perception*. New York: Routledge.

Carroll, N. (1996). Towards an ontology of the moving image. In *Theorizing the Moving Image*. New York: Cambridge University Press, 62–63.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36: 181–204.

——— (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Cohen, J. (2010). Perception and computation. *Philosophical Issues* 20: 96–124.

Cohen, J. and Meskin, A. (2004). On the epistemic value of photographs. *The Journal of Aesthetics and Art Criticism* 62: 197–210.

Cooper, M. (2008). *Line Drawing Interpretation*. London: Springer Verlag.

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.

Cutting, J. and Vishton, P. (1995). Perceiving layout and knowing distances. In W. Epstein and S. Rogers (eds.), *Perception of Space and Motion*. San Diego, CA: Academic Press, 69–117.

Dennett, D. (1969). *Content and Conscious*. London: Routledge and Kegan Paul.

Dretske, F. (1979). *Seeing and Knowing*. London: Routledge and Kegan Paul.

Firestone, C. (2013). How 'paternalistic' is spatial perception? Why wearing a heavy backpack doesn't—and couldn't—make hills look steeper. *Perspectives on Psychological Science* 8: 455–473.

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.

Fry, R. (1934). *Reflections on British Painting*. London: Faber.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gombrich, E. H. (1961/2000). *Art and Illusion*, 2nd revised edition. Princeton: Princeton University Press.

——— (1972). Illusion and art. In R. L. Gregory and E. H. Gombrich (eds.), *Illusion in Natural and Art*. New York: Charles Scribner's Sons, 193–243.

——— (1978). Illusion and visual deadlock. In *Meditations on A Hobby Horse, and Other Essays on the Theory of Art*. London: Phaidon, 151–159.

——— (1982). Mirror and map: Theories of pictorial representation. In *The Image and the Eye*: *Further Studies in the Psychology of Pictorial Representation*. Ithaca: Cornell/Phaidon Books, 172–214.

Hatfield, G. (1990). *The Natural and the Normative*. Cambridge, MA: MIT Press.

Hayek, F. (1952). *The Sensory Order*. Chicago: University of Chicago Press.

Hecht, H., Schwartz, R., and Atherton, M. (eds.). (2003). *Looking Into Pictures: An Interdisciplinary Approach to Pictorial Space*. Cambridge, MA: MIT Press.

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.

Hohwy, J., Roepstorff, A., and Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition* 108: 687–701.

Hopkins, R. (1998). *Picture, Image, and Experience*. Cambridge: Cambridge University Press.

——— (2006). The speaking image: Visual communication and the nature of depiction. In M. Kieran (ed.), *Contemporary Debates in Aesthetics and the Philosophy of Art*. Oxford: Blackwell, 135–159.

——— (2012a). Seeing-in and seeming to see. *Analysis* 72: 650–659.

——— (2012b). What Perky did not show. *Analysis* 72: 431–439.

Hopp, W. (2013). No such look: Problems with the dual content theory. *Phenomenology and the Cognitive Sciences* 12: 813–833.

Hume, D. (1740/1978). *A Treatise of Human Nature*. Oxford: Clarendon.

Kanizsa, G. (1979). *Organization in Vision: Essays on Gestalt Psychology*. New York: Praeger.

Kemp, M. (1989). *Leonardo on Painting: An Anthology of Writings by Leonardo da Vinci with a Selection of Documents Relating to His Career*. New Haven, CT: Yale University Press.

Kennedy, J. (1974). *A Psychology of Picture Perception*. San Francisco: Jossey-Bass Publishers.

Knill, D. and Richards, W. (eds.). (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Koenderink, J. J. (1998). Pictorial relief. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 356: 1071–1086.

——— (2012). *Pictorial Space*. Utrecht: De Clootcrans Press.

Koenderink, J. J., Van Doorn, A. J., and Kappers, A. M. (1994). On so-called paradoxical monocular stereoscopy. *Perception* 23: 583–583.

Koenderink, J. J., Van Doorn, A. J., Kappers, A. M., and Todd, J. T. (2004). Pointing out of the picture. *Perception* 33: 513–530.

Koenderink, J. J., Wijntjes, M., and van Doorn, A. J. (2013). "Zograscopic Viewing." *i-Perception* 4: 192–206.

Köhler, W. (1913/1971). On unnoticed sensations and errors of judgment. In M. Henle (ed.), *The Selected Papers of Wolfgang Köhler*. New York: Liveright, 13–39.

Koslow, S. (1967). De wonderlijke perspectyfkas: An aspect of seventeenth century Dutch painting. *Oud Holland* 82: 35–56.

Kubovy, M. (1986). *The Psychology of Perspective and Renaissance Art*. Cambridge: Cambridge University Press.

Kulvicki, J. (2009). Heavenly sight and the nature of seeing-in. *Journal of Aesthetics and Art Criticism* 67: 387–398.

Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research* 35: 389–412.

Leopold, D. A. and Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences* 3: 254–264.

London, J. (1909/1982). Martin Eden. In *Novels and Social Writings*. New York: Library of America.

Lopes, D. (1996). *Understanding Pictures*. Oxford: Clarendon Press.

——— (2005). *Sight and Sensibility: Evaluating Pictures*. Oxford: Oxford University Press.

——— (2010). Picture this: Image-based demonstratives. In Catharine Abell and Katerina Bantinaki (eds.), *Philosophical Perspectives on Depiction*. Oxford: Oxford University Press, 52–80.

Lycan, W. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

——— (2008). More layers of perceptual content. Presentation at *Naturalized Philosophy of Mind and Language: A Conference in Honor of Ruth Garrett Millikan*, October 4, 2008, University of Connecticut.

Macpherson, F. (2015). The structure of experience, the nature of the visual, and type 2 blindsight. *Consciousness and Cognition* 32: 104–128.

Mamassian, P., Landy, M., and Maloney, L. (2002). Bayesian modeling of visual perception. In R. Rao, B. Olshausen, and M. Lewicki (eds.), *Probabilistic Models of the Brain*. Cambridge, MA: MIT Press, 13–36.

Martin, M. G. F. (2012). "Sounds and Images." *British Journal of Aesthetics* 52: 331–51.

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

McGinn, C. (2004). *Mindsight: Image, Dream, Meaning*. Cambridge, MA: Harvard University Press.

——— (2005). *The Power of Movies*. New York: Pantheon.

Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics* 13: 59–66.

——— (1974). The perception of transparency. *Scientific American* 230: 90–98.

Michotte, A. (1960/1991). The real and the unreal in the image. In G. Thines, A. Costall, and G. Butterworth (eds.), *Michotte's Experimental Phenomenology of Perception*. Hillsdale, NJ: Erlbaum, 187–196.

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

Nakayama, K., He, Z., and Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. Kosslyn and D. Osherson (eds.), *Visual Cognition*. Cambridge, MA: MIT Press.

Nanay, B. (2010). Inflected and uninflected perception of pictures. In C. Abell and K. Bantinaki (eds.), *Philosophical Perspectives on Pictures*. Oxford: Oxford University Press, 181–207.

——— (2014). *Trompe l'oeil* and the dorsal/ventral account of picture perception. *Review of Philosophy and Psychology* 6: 181–197.

Newall, M. (2015). Is seeing-in a transparency effect? *The British Journal of Aesthetics 55*: 131–156.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Norman, L. J., Akins, K., Heywood, C. A., & Kentridge, R. W. (2014). Color constancy for an unseen surface. *Current Biology* 24(23): 2822–2826.

——— (2005). Real presence. *Philosophical Topics* 33: 235–264.

Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

Peterson, M., Gillam, B., and Sedgwick, H. A. (eds.). (2007). *In the Mind's Eye: Julian Hochberg on the Perception of Pictures, Films, and the World*. Oxford: Oxford University Press.

Pirenne, M. H. (1970). *Optics, Painting, and Photography*. London: Cambridge University Press.

Prinz, J. (2006). Beyond appearances: The content of sensation and perception. In T. Gendler and J. Hawthorne (eds.), *Perceptual Experience*. Oxford: Clarendon Press.

Purves, D., Lotto, B., and Nundy, S. (2002). Why we see what we do. *American Scientist* 90: 236–243.

Rescorla, M. (2015). Bayesian perceptual psychology. In M. Matthen (ed.), *The Oxford Handbook of the Philosophy of Perception*. Oxford: Oxford University Press.

Rogers, S. (1995). Perceiving pictorial space. In W. Epstein and S. Rogers (eds.), *Perception of Space and Motion*. London: Academic Press Limited, 119–164.

——— (2003). Truth and meaning in pictorial space. In H. Hecht, R. Schwartz, and M. Atherton (eds.), *Looking Into Pictures: An Interdisciplinary Approach to Pictorial Space*. Cambridge, MA: MIT Press, 301–320.

Ruskin, J. (1856/1971). *The Elements of Drawing*. New York: Dover.

Sartre, J. P. (2004). *The Imaginary: A Phenomenological Psychology of the Imagination* (trans. J. Webber). London: Routledge.

Schlosberg, H. (1941). "Stereoscopic Depth from Single Pictures." *American Journal of Psychology* 54: 601–5.

Schwartz, R. (2006). *Visual Versions*. Cambridge, MA: MIT Press.

Schwitzgebel, E. (2006). Do things look flat? *Philosophy & Phenomenological Research* 72: 589–599.

——— (2008). The unreliability of naive introspection. *Philosophical Review* 117: 45–273.

Sedgwick, H. (1986). Space perception. In K. L. Boff, L. Kaufman, and J. P. Thomas (eds.), *Handbook of Perception and Human Performance, Vol. 1: Sensory Processes and Perception*. New York: Wiley.

Siegel, S. (2010). *The Contents of Visual Experience*. New York: Oxford University Press.

——— (2016). Rich or thin? In B. Nanay (ed.), *Contemporary Debates in the Philosophy of Perception*. New York: Routledge.

Siewert, C. (2006). Is the appearance of shape protean? *Psyche* 12(3): 1–16.

Smith, A. D. (2000). Space and sight. *Mind* 109: 481–518.

Snowdon, P. (2015). Sense-data. In M. Matthen (ed.), *The Oxford Handbook of the Philosophy of Perception*. Oxford: Oxford University Press.

Thompson, W., Fleming, R., Creem-Regehr, S., and Stefanucci, J. K. (2011). *Visual Perception From a Computer Graphics Perspective*. New York and London: CRC Press.

Tolman, E. C. and Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review* 42: 43–77.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology* 6: 171–178.

Tye, M. (2003). Blurry images, double vision, and other oddities: New problems for representationalism? In Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Essays*. Oxford: Oxford University Press, 7–32.

Vishwanath, D. (2014). Toward a new theory of stereopsis. *Psychological Review* 121: 151–178.

Vishwanath, D. and Hibbard, P. (2013). Seeing in 3-D with just one eye: Stereopsis without binocular vision. *Psychological Science* 24: 1673–1685.

Walton, K. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press.

——— (2008). *Marvelous Images: On Values and the Arts*. Oxford: Oxford University Press.

White, J. (1967). *The Birth and Rebirth of Pictorial Space*. London: Faber and Faber.

Wiesing, L. (1967). *Zettel* (trans. G. E. M. Anscombe and G. H. von Wright). Oxford: Blackwell.

——— (1980). *Remarks on the Philosophy of Psychology*, Vol. II. Oxford: Blackwell.

——— (2010). *Artificial Presence: Philosophical Studies in Image Theory*. Stanford: Stanford University Press.

Wölfflin, H. (1929). *Principles of Art History* (trans. M. D. Hottinger). New York: Dover Publications.

Wollheim, R. (1963). Review of *Art and Illusion*. *The British Journal of Aesthetics* 3: 15–37.

——— (1974). On drawing an object. In *On Art and the Mind*. Cambridge, MA: Harvard University Press, 3–30.

——— (1980). Seeing-as, seeing-in, and pictorial representation. In *Art and Its Objects*. Cambridge: Cambridge University Press, 205–226.

——— (1987). *Painting as an Art*. London: Thames & Hudson.

——— (1993). Pictures and language. In *The Mind and Its Depths*. Cambridge, MA: Harvard University Press, 185–192.

——— (1998). On pictorial representation. *The Journal of Aesthetics and Art Criticism* 56: 217–226.

——— (2003). What makes representational painting truly visual? *Aristotelian Society Supplementary* 77: 131–147.

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, 298–328.