

**John Zeimbekis**

**Seeing, visualizing, and believing: Pictures and cognitive penetration**

Visualizing and mental imagery are thought to be cognitive states by all sides of the imagery debate (Tye, Pylyshyn and Kosslyn). Yet the phenomenology of those states has distinctly visual ingredients. This has potential consequences for the hypothesis that vision is cognitively impenetrable, the ability of visual processes to ground perceptual warrant and justification, and the distinction between cognitive and perceptual phenomenology. I explore those consequences by describing two forms of visual ambiguity that involve visualizing: the ability to visually experience a picture surface as flat after it has caused volumetric nonconceptual contents (§§2-3), and the ability to use a surface initially perceived as flat to visualize three-dimensional scenes (§4). In both cases, the visual processes which extract viewer-centered volumetric shapes (equivalent to Marr's 2½D sketch) have to rely solely on monocular depth cues in the absence of parallax and stereopsis. Those processes can be cognitively penetrated by acts of visualizing, including ones that draw on conceptual information about kinds. However, the penetrability of the visual processes does not weaken their ability to provide perceptual warrant and justification for beliefs (§5). The reason is that picture perceptions—whether they are stimulus-driven or based on acts of visualizing—are different to object perceptions both phenomenologically and in terms of their functional roles as states. Thus, although the penetrability of the visual processes does mean that subjects can have visual experiences with contradictory (2D and 2½D) contents, perceptual belief is adopted at most towards

one set of contents, and questions of warrant and justification are raised only for those contents. A rule-proving exception is provided by *trompe-l'oeils* (§6).

I use the expressions '*volumetric content*', '*2D contents*' and so forth to designate the representational content of visual experiences, whether the experiences are caused by object perceptions or by picture perceptions, and whether they are stimulus-driven or driven jointly by acts of visualizing. I take such visual experiences to be representational, contentful states which divulge accuracy or correctness conditions. The accuracy conditions can be thought of in terms of Peacocke's (1992) scenario content. For example, to have a visual experience with volumetric contents is to represent volumetric shapes at certain egocentric locations. What it is to have visual experiences with volumetric representational contents when perceiving flat surfaces, and what kind of epistemic predicament this is, are explained at length below. On the other hand, why visual experiences have content and accuracy conditions in the first place, and what it is to have content and nonconceptual content in particular, are not topics dealt with in this paper. For concise replies to the last two questions, see Macpherson, this volume, §1.

## **I. Pictures and visual ambiguity**

Among theories of depiction, there is considerable consensus that pictures can cause not only visual experiences with volumetric contents but also veridical visual experiences of picture surfaces as flat. The assumption is so widespread that it splits into several varieties, depending on whether viewers can have both visual experiences at once. One variety holds

that it is *possible* to have both the experience of volumetric shapes and the experience of flatness at once (Peacocke 1987, 386, 394; Schier 1986, 9; Lopes 1996, 40; Walton 1990, 293-304). Another says that we *always* have both experiences at once (Wollheim 1987, 46; Gibson 1986, 282; Rock 2001, 98). A third holds that we *cannot* have both experiences at once, but can have each one separately (Gombrich 1960, 224, and perhaps Cutting and Massironi, 1998). But there is one point on which all of these authors agree: that we are capable of having both kinds of visual experience, both of flatness and of depth, of picture surfaces. If this consensus on the phenomenology of picture perception is anything to go by, pictures qualify as *visually ambiguous* figures. The consensus is worth noting because, as Macpherson has said in a related context,<sup>1</sup> ‘the evidence that one must appeal to here is introspective and the more people that agree that a change takes place in their experience the better’ (Macpherson 2006, 91).

Yet, if we admit the phenomenological claims as a form of evidence, then pictures should support not one but *several* kinds of visual ambiguity. In one kind, visually experiencing a picture (or even parts of it at a time) as a flat surface (one perpendicular to the line of vision) requires great mental effort. Figure 1 provides an illustration. If we can visually experience such pictures or parts of them at a time as flat, it will only be *subsequent* to the pictures’ having caused visual experiences with volumetric content; the mental effort could involve consciously shifting attention across the surface until we find a way to blind

---

<sup>1</sup> The context is Macpherson’s discussion of the ability to perform Gestalt switches with ambiguous figures; but in the cases Macpherson has in mind the switch is not from 3D (2½D, ie, egocentrically 3D) to 2D visual contents, as will be the case with the examples introduced here.

ourselves to regions that work as depth cues and prevent the cues from triggering depth representations. Even Peacocke's relatively modest claim, that we experience the outlines of object-representations in pictures as flat,<sup>2</sup> seems wrong where this picture is concerned: the left outline edges of the pyramids seem further away than the right edges. The problem is that vision refuses to process *just* the outlines; it processes the interior lines as edges and yields a mental representation of a regular volume with a certain orientation, which settles the egocentric distances of the object's parts, including parts of the outlines.



*Figure 1*

Whether or not figure 1 is in fact visually ambiguous (as one would expect given the claims that pictures can also cause visual experiences of their surfaces as flat), it is clear that when we perceive it we *first* become aware of volumetric visual contents and *then* have trouble suppressing this visual experience by making an agentive mental effort. How

---

<sup>2</sup>The silhouette both is and is experienced as a flat surface, or at least as occupying a plane: any description of the experience omitting this point is incomplete.' Peacocke 1987: 386.

do the volumetric visual contents of such ‘natural’ picture perceptions emerge in the first place? We could sketch the following hypothesis, postponing until the next section its details and justification. The picture’s surface includes monocular depth cues which can trigger brain representations of volume even without calculations from parallax and stereopsis; those cues exploit early visual processes dedicated to constructing volumetric contents; the outputs of those early processes are brain representations corresponding to Marr’s 2½D sketch; those representations provide the contents of phenomenal awareness; and those phenomenal contents constitute the visual experience of the picture’s content. If the hypothesis is plausible, as I will try to show in §2, then, in this form of visual ambiguity, volumetric contents are experienced without any help from visualizing or visually imagining, which can be driven consciously and are usually thought to be higher-level processes. (See Tye 1991, 90, 96; Pylyshyn 2007, 141; Pylyshyn 2003; Kosslyn 1994.) On the other hand, 2D contents emerge—if they emerge at all, or when they do<sup>3</sup>—only subsequently, as a result of some conscious activity.

Today we frequently have such picture perceptions, which feel effortless and automatic, because many of the pictures we perceive (including non-agentive pictures like photographs) contain cues which successfully exploit object-perception processes like segmentation and the construction of egocentric depth representations out of two-dimensional distributions of light. But while some depth cues (line junctions) were mastered in prehistoric times (Cavanagh 2005, Biederman and Kim 2008), it took a long time for drawing and painting to develop techniques which could cause visual and

---

<sup>3</sup>I think they do emerge in some cases; see §3.

perceptual contents comparable to those caused by objects about shading, textures, colours, lines, orientations, volumes, and the depth relations between objects (see, for example, Kubovy 1986 on depth relations; Gombrich 1960, Ch. 1, on colour). Moreover, once mastered, the techniques were often subsequently avoided by painters trying to develop new pictorial styles—that is, new ways of causing a visual impression of depth and volume without relying on existing techniques. When a picture does not include adequate cues to evoke some of those features (shading, textures, lines, orientations, etc), the brain cannot form representations of the intended picture contents solely on the strength of stimulus-driven or hard-wired processes. On the strength of those processes, the brain may represent the picture as a flat surface with uneven stains, or else it may capture only some local depth features (like apparent occlusions) but not global ones or depth relations between objects. In those cases, we have to perform conscious acts of visualizing by using the picture surface as a prop.

This gives rise to a second kind of visual ambiguity supported by pictures. This time, the surface naturally yields a visual experience of a flat surface; but subsequently, by performing conscious acts of visualizing, we can look at the surface and generate experiences of depth and volume. Such cases of depiction are often provided by what Cutting and Massironi (1998) call *fortuitous* pictures: natural or other objects whose colour and line patterns can, but do not always, support visualizing activities which allow ‘seeing-in’. The passage below, taken from Leonardo’s advice to apprentice painters, seems to describe this kind of ambiguity:

if you look at any walls soiled with a variety of stains, or stones with variegated patterns, when you have to invent some location, you will therein be able to see a resemblance to various landscapes graced with mountains, rivers, rocks, trees, plains, great valleys and hills in any combinations. Or again you will be able to see various battles and figures darting about, strange-looking faces and costumes, and an endless number of things which you can distill into finely-rendered forms. And what happens with regard to such walls and variegated stones is just as with the sound of bells, in whose peal you will find any name or word you care to imagine. (Leonardo 1989, 222)

For instance, it is possible to visually experience figure 2 as the stained flat surface that it is. But it is also possible to visually experience the figure depthwise as figure and ground; and in fact concepts can help us to settle the figure-ground relation, for instance if we're told to see the lower part as a field of wheat and the upper part as a grey sky. A question that emerges about such cognitively driven acts of visualizing is whether they just add a cognitive ingredient to mental contents without altering the outputs of the visual processes leading up to Marr's 2½D sketch, or instead tamper with the way vision itself generates figure-ground segregations; I postpone this discussion until §4.



*Figure 2*  
Turner, *Approach to Venice*, oil on canvas, 1844; detail.  
Courtesy National Gallery of Art, Washington.

A third form of picture-related visual ambiguity occurs when neither 2D nor 2½D visual contents are the more natural output of vision, so that neither prevails in perception. Details of pictures taken out of context can be ambiguous in this way because their depth cues can easily be overridden in isolation. Many Rorschach blots are also in this category: they cause representations of volume only indirectly (through recognition, by exploiting templates for outlines); they do not include shading and enclosed lines that could signal convexities, concavities or what Marr calls surface orientation discontinuities (Marr 1982, 215-233). The weakness of Rorschach blots as pictures is described by Gibson when he writes that their ‘invariants are all mixed up together and are mutually discrepant instead of being mutually consistent or redundant’ (Gibson 1986, 282). A fourth kind of ambiguity is bi-stability in which both visual interpretations are volumetric and both are generated bottom-up, as apparently occurs when we switch among different perceptions of the Necker cube. It seems plausible to hold, as Pylyshyn does (2003, Ch. 1), that both visual experiences of the Necker cube are performed purely by bottom-up processes jointly with



hard-wired processes. Finally, a fifth kind of ambiguity is that found in perceptions of Jastrow's duck-rabbit drawing. Both experiences caused by the drawing are experiences of the picture's content, namely, an object belonging to a three-dimensional kind. However, if those experiences involve any representation of volume, that representation depends almost entirely on prior knowledge encoded in sortal concepts, not stimulus-driven depth cues. In fact, it is not immediately clear that there is any change of *visual* contents in the Jastrow's ambiguity, although this has been claimed recently (see Macpherson 2006, 97; Siegel 2010 makes a related point, discussed here in §5). In the Jastrow figure, representations of volume and depth are not caused by bottom-up or hard-wired visual processes, as they are by the Necker cube, because the Jastrow contains no internal lines suggestive of surface discontinuities—only a couple of weak volumetric cues from shading (the shading that suggests a concavity between the head and neck). So any change of visual experience as we switch from one experience to the other is unlikely to be a change in *content* construed as what would have to be the case in the world for the experience to be true. The change seems to be more superficial; it has been described by Lyons as 'manipulating the representation produced' previously by the bottom-up processes without altering it, or as 'facilitating pop-out of certain patterns' and yielding 'a late experiential effect, leaving the nonconceptual early perceptual states unaffected but influencing the nondoxastic seemings' (Lyons 2011).

While cases belonging to all five kinds of ambiguity have been described and discussed in the literature on depiction, the kinds of visual ambiguity supported by pictures have not yet been distinguished, nor the differences between them explained. Making the distinctions is

important for two sorts of reasons. First, ambiguous figures have always been used to test claims about the cognitive impenetrability of certain visual processes and certain ways of drawing the perception-cognition distinction (Churchland 1988, Macpherson 2006). Deniers of cognitive penetrability (Fodor 1988, Pylyshyn 1999, Raftopoulos 2009, 2011) have in turn responded to each kind of ambiguity-based penetrability claim. Some kinds of ambiguity are easy enough for modularists to deal with, others less so. But pictures are an extremely rich resource of different kinds of visual ambiguity; the first two kinds isolated above have not been described in the penetrability literature but pose a significant threat to the impenetrability hypothesis. Secondly, understanding the different forms of visual ambiguity is essential to understanding what pictures are. For example, they show that picture perception is not a single kind of mental state, be it a higher-level state of visualizing or a lower-level state which is the output of earlier visual processes;<sup>4</sup> they allow us to describe the mental states and contents caused by different kinds of pictures, and thus to relate those states and contents to other states such as perceptual belief and memory; they allow us to account for what is called naturalism in depiction; and so on. Thus, issues about the perception-cognition distinction turn out to be essential for an account of depiction; though in this paper I will focus on the first set of issues.

---

<sup>4</sup>For example, Abell and Currie (1999, 440), Wollheim (1987, 1998) and Levinson (1998, 232) seem to be committed to an account of picture perception as the output of pre-doxastic processes; Walton (1990) to a doxastic-level account (though see Walton 2002, 31, for a denial of this). Most authors on depiction do not address the issue at all; an exception is Levinson 1998.

The first two kinds of visual ambiguity described correspond to two kinds of picture perception. The first, illustrated by figure 1, could be called bottom-up or *natural picture perception*. The ambiguity it gives rise to is one in which an object naturally causes 2½D visual contents, but with some mental effort we can also see the object as flat. The second, illustrated by figure 2, could be called *cognitively driven picture perception*; the ambiguity it gives rise to is one in which an object naturally causes 2D visual contents, but with some mental effort we can see the object as 2½D.

How exactly do these two kinds of visual ambiguity relate to the cognition-perception distinction and to cognitive penetrability? If all picture perceptions were bottom-up and hard-wired, or if all picture perceptions were cognitively driven acts of imagining, or if some were bottom-up and others cognitively driven, this would not necessarily entail any form of cognitive penetrability. The bottom-up picture perceptions could be compatible with impenetrability of the visual processes up to Marr's 2½D sketch, while the cognitively driven ones could conceivably be such that they added a cognitive phenomenology to mental contents without altering the outputs of the visual processes leading up to Marr's 2½D sketch. It is along similar lines that Raftopoulos (2009, 2011), Fodor (1988) and Pylyshyn (2003) have dealt with the kinds of ambiguity supported by the Jastrow figure and the Necker cube. But there is a difference between those forms of visual ambiguity and the forms described above which involve switching from 2D to 2½D visual contents and vice versa: only the latter two have the potential to cause trouble for the hypothesis that visual processes required to construct 2½D nonconceptual content out of the 2D sketch are cognitively impenetrable.

My strategy in this paper is twofold. I will concede that there are good counterexamples to the claim that visual processes leading up to Marr's 2½D sketch are cognitively impenetrable. Sections 3 and 4 describe two forms of visual ambiguity and argue that they imply the cognitively penetrability of visual processes required to construct 2½D content out of the 2D sketch. I believe that trying to exclude all such cases of penetrability would be barking up the wrong tree. Instead—and this is the second part of the strategy—I will claim that even if it turns out that the functioning of some key visual processes can sometimes be tampered with by processes that count as cognitive, this does not lead to the expected pernicious epistemic and epistemological consequences (sections 5 and 6). The reason is that visual experiences caused by pictures fail to cause perceptual beliefs. This also amounts to something of an enabling condition for depiction as a form of representation, as opposed to a source of illusions.

## **2. Bottom-up or natural picture perceptions**

The hypothesis that some pictures cause visual experiences with 2½D contents without any conscious mental contribution on behalf of the viewer is plausible, because otherwise picture perception would always be an effortful, slow performance for subjects, involving acts of visualizing for each feature of the picture that can be used to imagine a depth relation or a volumetric shape. Instead, most picture perception is caused in an effortless, quick and automatic way, suggesting that it is subserved by hard-wired and bottom-up processes, not deliberate acts of imagining. So the hypothesis is compatible with the

phenomenology of many picture perceptions; and it also allows us to explain the other cases: in the absence of cues that trigger automatic, dedicated processes, we have to perform slow, attention-consuming acts of visualizing to grasp the picture contents. A sketch follows of how bottom-up picture perceptions could occur. It focuses mainly on the construction of volumetric shapes from monocular cues, and for reasons of space omits the segregation of objects during picture perception and the ways in which different object-recognition processes are co-opted by different kinds of pictures (issues which are discussed in Zeimbekis 2012).

The natural place to look for support for the claim that many pictures cause egocentric, conscious representations of 3D shapes is Marr's theory of vision. Marr's working hypothesis is that initial, two-dimensional retinal inputs are built up into adequate representations of the three-dimensional objects that cause them, and this makes its details particularly interesting for a theory of depiction. An important part of Marr's proposal cannot be applied to pictures. It is the part of his theory (Marr 1982, Ch. 3) according to which binocular disparity, the slight difference in perspective from which each eye views the object, is exploited by the visual system to compute the relative depths of surfaces in a scene. In pictures, all the points of the surface from which the eyes receive stimuli are at roughly the same location on the back-to-front axis (the axis perpendicular to the picture's surface and passing through the centre of the viewer's body, which is identical to the  $z$  axis of Peacocke's nonconceptual scenario content; see Peacocke 1992, 62). Therefore, in picture perception, the brain cannot compute depth relations from binocular disparity and has to rely on monocular depth cues. The same applies to parallax from movement relative

to the object. When we move relative to the picture, points on the surface are not seen as moving at different speeds and parallax cannot be exploited to compute depth. Picture perception is even insensitive to changes in the angle from which the picture is viewed; the picture can be rotated up to 22 degrees on its vertical axis without this deforming the three-dimensional representations it causes (Cutting 1987), an effect Kubovy (1986) calls 'robustness of perspective'.

However, Marr also relies heavily on hypotheses about how we extract volumetric and depth information from *monocular cues* which exploit neither binocular disparity nor parallax. Here, his proposal is that parts of the visual system take two-dimensional representations of objects or scenes as inputs, apply hard-wired processes to them, and yield three-dimensional shape representations. The inputs of the processes are edges (lines) and their orientations and junctions, that is, the outputs of earlier visual processes which generate a brain representation of differences in light intensities. These are treated by vision as discontinuities in the distance of surfaces from the viewer ('occluding contours'), convexities or concavities of surfaces ('surface orientation discontinuities') and depthwise curvature ('surface contours') (1982, 215-233). Texture patterns are also interpreted to yield representations of the depthwise slant of a surface (233-239). According to Marr, the brain applies a set of simplifying constraints when interpreting occluding contours; for example, points close on the 2D contour are assumed to be close in 3D space (223), an assumption which is nicely illustrated by Richard Gregory's impossible triangle. When those assumptions are applied to two-dimensional views of certain regular 3D shapes,<sup>5</sup>

---

<sup>5</sup>The regular shapes are 'generalized cones', shapes whose cross sections have the same shape but can vary in size (for example, a sphere, a pyramid or a cylinder). Although the objects we have to recognize in the

they yield accurate three-dimensional representations of those shapes. (Note that monocular cues are likely to be an essential resource for vision when objects are too distant for binocularity or parallax to calculate the relative distances of parts of an object.)

On Marr's model, the output of these processes—the 2½D sketch, an egocentric volumetric representation—subserves kind-recognition and therefore precedes the application of concepts. Volume is assigned as a condition for recognition and classification. The opposite—classification as a condition for assigning shape—would amount to a form of sortalism like the one criticized by Campbell (see his account of the 'delineation thesis'; Campbell 2002, 69). On the sortalist scenario, we would carve up the visual scene into volumetric objects by using sortal concepts, not on the basis of bottom-up stimuli jointly with hard-wired visual processes like those that Marr describes; those processes would under-determine the assignment of shape, allowing different shapes to be assigned depending on which concepts we applied.

A sceptic might argue that in picture perception, due to the absence of any differential stimuli from stereopsis and without performing the processes that compute them, the brain would not be in the kinds of states that qualify as causal antecedents for having three-dimensional object representations. While we do have 3D object representations at *some* point in picture perception, the sceptic might argue that they are not caused by lower-level

---

visual environment are not usually generalized cones, they can be analysed coarsely into regular component shapes for the purposes of kind-recognition (Marr 1982, Biederman 1987, 1995; see also the brief discussion below).

visual processes but are instead the effects of some kind of higher-order imagining or visualizing. My response to this scepticism is twofold. First, I do not deny that there are cognitively driven picture perceptions which involve conscious acts of imagining. Such picture perceptions are different to the ones I call ‘bottom-up’ or ‘natural’, and the sceptic would have to explain the differences. Admitting bottom-up picture perceptions can explain the difference between perceiving pictures like Figure 1 and perceiving pictures in non-naturalistic styles like cubism. For example, Picasso’s *The guitar player* (1910) can support a visual representation of a guitar player—at least if the concept of visual representation is construed to include the contents of visualizing. But to succeed in having the visual representation of a human body in the position of a guitar player, the subject first has to make conscious hypotheses about depth relations and volumetric shape. Before we read the picture’s title and make use of concepts, all we can pick up in terms of volume or depth are, at best, some local occlusion effects which do not allow us to reconstruct the volumes corresponding to the object described by the concepts. Figure 1 on the other hand yields that kind of information naturally and effortlessly as soon as we glance at the picture, suggesting that Marr is right to hold that the brain can reach volumetric object representations without either the benefit of binocular disparity or input from recognitional concepts.

Object-recognition theories after Marr seem to confirm this hypothesis. According to one such theory, Biederman’s theory of recognition by components (‘RBC’; Biederman 1987, 1995), the features of objects on the basis of which we analyse them into components, such as vertices, which are interpreted as concavities or convexities, are ‘generally invariant



over viewing position' (1987, 115). As a result, the information required to analyse an object into primitive volumetric components can be extracted from a single two-dimensional representation of the object (1995, 153; 1987, 133-141). A picture of an object causes a two-dimensional view of that object from a single viewpoint, so on this theory, picture perceptions can suffice to cause volumetric object representations. Note that on RBC theory, these structural representations are object centered, not viewpoint relative; they correspond to the objects in Marr's 3D sketch of the visual scene and are thought to be the immediate causal antecedents of object recognition on this theory as well as Marr's.

Competing views of object recognition have shown that RBC is inadequate for explaining certain recognition tasks and fails to match a quantity of experimental data. The competing accounts are called 'image-based', 'viewpoint-relative' or 'multiple-views' theories. Image-based or viewpoint-relative models of object recognition emerged from evidence that recognition also relies on viewpoint-relative information in ways that RBC does not account for (Tarr and Pinker 1989, Bulthoff and Edelman 1992, Edelman and Bulthoff 1992, Tarr 1995, Ullman 1998). In a series of experiments, subjects who were acquainted with new kinds of objects could still only recognize them from viewpoints similar to those under which they were presented to them, contradicting the predictions of RBC theory. (This behavioural evidence has also been backed by neurological findings; see Logothetis et al. 1995.) Moreover, RBC's structural representations are too coarse-grained to capture the recognition of individuals, or even to subserve many fine-grained generic classifications. These theories could be seen to present a challenge to the idea that a volumetric (egocentric, 2½D) representation is constructed prior to recognition, since they

could be compatible with the claim that volumetric representation is added cognitively post-recognition, not visually prior to recognition. However, viewpoint-dependent theories do *not* make the claim that the brain representations that trigger recognition are representations of objects in *two* dimensions. The relevant distinction is that between structural, object-centered representations, and representations that are relativized to *viewpoints*. The representations in Marr's 2½D sketch are also relativized to viewpoints, yet they are volumetric and represent depth; they contrast with subsequent brain representations which permit mental rotation, belong to the 3D sketch, and qualify as structural or viewpoint-*independent*. The conclusion that both viewpoint-relative and viewpoint-independent mechanisms are involved in recognition is supported by neurological findings on an area of the human brain called the lateral occipital complex (LOC). According to Kourtzi et al. (2003), this area represents perceived object-shape, a relatively high-level cognitive representation, and plays a role in object recognition. Kourtzi et al. show that while one subregion of the LOC represents two-dimensional shape, another subregion encodes the represented three-dimensional shape of objects and appears to 'mediate object and scene recognition based on rather abstract three-dimensional representations' (Kourtzi et al., 2003, 918; see 911 for other studies which show that the LOC represents shape three-dimensionally).

Volumetric representations on Marr's hypothesis and on the RBC hypothesis, and viewpoint-relative representations on the Tarr's, Bulthoff's and Ulmann's hypotheses, are considered causal antecedents for object recognition and classification, and therefore must be *nonconceptual* states. Putting aside for a moment the non-naturalistic and fortuitous

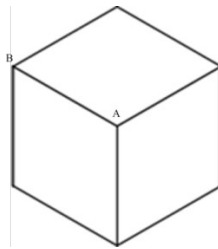
pictures that require agentively driven attention and conceptual input to contribute to depth interpretation, and focusing for on the important category of naturalistic pictures, this yields the following result. Descriptions of the nonconceptual contents of perception by Evans (1982), Peacocke (1992), and recently Raftopoulos (2009), are consistent with the descriptions Marr gives of the 2½D sketch. Both are perception-dependent, unlike conceptually encoded representations; both are spatially egocentric; and in both cases, shape, orientation, texture and colour (including shading) information is more fine-grained than prior conceptually encoded information about such properties. (For examples, see Marr's illustrations of how we visually experience flat surfaces that contain monocular depth cues; 1982, 215-239.) On several current theories of the neural correlates of consciousness, the kinds of shape representations which according to Marr constitute the 2½D sketch require local recurrent processing—a kind of brain state that occurs at approximately 100 to 120 milliseconds after stimulus onset and precedes personal-level awareness—which is thought to be the neural correlate of phenomenal awareness in visual perception (Block 1990, 1997, 2005; Lamme 2000, 2003; Raftopoulos 2009, 32-39). If that hypothesis is true, then the egocentric volumetric shape representations of Marr's 2½D sketch constitute the nonconceptual contents of visual perception.

Thus, the sketch of bottom-up picture-perceptions that I have drawn here would explain an important part of the phenomenology of picture perception: by the time we reach personal-level awareness and the ability to think about the picture, the picture has already caused a visual experience of a three-dimensional scene. This conclusion seems accurate phenomenologically; for a great number of pictures, picture perception is not the result of

any mental effort but occurs naturally. However, the bottom-up account of many picture perceptions also raises new questions: if there are bottom-up picture perceptions, then what differentiates those experiences from perceptual illusions, and especially from the visual experiences caused by trompe l'oeils? These questions are dealt with in §6.

### **3. Picture perception and cognitive penetrability**

As stated in §1, there is considerable consensus among depiction theorists that picture perceptions can support both 2½D and 2D visual contents. An easy way to account for that insistence would be to hold that such claims of visual ambiguity are limited to cognitively driven picture perceptions and do not apply to bottom-up picture perceptions. This response would sidestep the underlying issue of how viewers could avoid having contents caused by bottom-up processes. The problem with the response is that the theories tend to claim that pictures *generally*—not just pictures that require cognitively driven forms of picture perception—can support such visual ambiguity, and rely on naturalistic pictures such as Constable's paintings. Consider figure 3, which is similar to one of the drawings Marr (1982, 221) uses to illustrate the interpretation of surface orientations as volumetric shapes by early, hard-wired visual processes. The natural visual interpretation of the figure yields a nonconceptual representation of a cube seen from above, with point A in front of point B.



*Figure 3*

Marr's explanation of the experience of depth is this: 'If the occluding contour shown with thick lines is present on its own, one perceives a hexagon. The interior lines change it into a cube, since they suggest that the occluding contour is not planar' (1982, 221). But although it is initially *difficult* to perceive this figure and visually represent it as flat, it is *possible*. It suffices to be told to look for three rhomboids, or for a regular hexagon, or to see the enclosed lines as radii. If you see the hexagon, attention is distracted from point A to the perimeter and you experience a short-lived change of visual experience. Note that figure 3 is not bistable: we return naturally to the cube-perception once we relax the effort of attention, so there is one visual interpretation we could call *natural*. Natural picture perceptions share this characteristic with what Kanizsa and Gerbino call 'amodal completions', as distinct from what they call 'represented' or 'merely thought' completions, which work like cognitively driven picture perceptions. (For examples of such revisable amodal completions see Kanizsa and Gerbino, 1982, Figures 9.3.c and 9.6.a.)

There appear to be two different ways to go about seeing the drawing as flat. The first is to consciously direct attention away from the drawing's enclosed lines and on to the perimeter, optionally by using the concept *rhomboid* or *hexagon* to focus attention differently. That way, we can avoid performing the hard-wired interpretation of the intersection of the three enclosed lines as three-dimensional. Conscious manipulations of attention are a personal-level activity even when they are not concept-driven; when they are concept-driven (eg because we are given *hexagon* or *rhomboid*), the concept determines where attention is directed.

But there is a second way to get the same effect, which this time does *not* require us to attentionally ignore the enclosed lines in Marr's cube: we can just see them differently, as radiating from the centre of the hexagon. In that case, we focus on the very part of the picture that contains its depth cue and *still* get a 2D representation. If that is so, and if we appeal to the attention-shift argument to explain the change in visual experience and content, we're exposed to the rejoinder that viewers *directly* interfere with the processes that construct volumetric representations. The rejoinder may or may not be correct, but as it stands it pits one introspective claim against another. If the dispute between deniers and defenders of penetrability hinges on such a conflict, then it seems to reach a stalemate. This is important because the key argument by deniers of penetrability where ambiguous figures are concerned is that shifts in where spatial attention is focused on the scene change *which* data is processed, without changing *the way* such data is processed (Fodor 1988; Pylyshyn 1999; Raftopoulos 2009, 2011). If Marr's opaque cube is visually ambiguous in the way described, then it is a clear counterexample to the attention-shift

argument for a key part of perception: the construction of 2½D visual contents from lines of the optic array that signal surface orientation discontinuities and convexities.

Which other strategies are available to block the conclusion that the switch from 2½D to 2D visual contents implies the penetrability of visual processes by agentively driven processes? For one kind of ambiguity seen in §1, that found in Jastrow's rabbit-duck figure, it can be argued that what changes each time is only the *conceptual* representation caused by the nonconceptual visual contents, and that the visual experience and contents remain the same throughout. That argument has been used to counter Churchland's use of perceptual ambiguity to challenge Fodor's (1983) claims about the impenetrability of visual processing modules. Churchland writes that we can make figures like the Jastrow 'flip back and forth at will between the two or more alternatives, by changing one's assumptions about the nature of the object or about the conditions of viewing,' concluding that 'some aspects of visual processing, evidently, are quite easily controlled by the higher cognitive centers' (Churchland 1988, 172). Discussion of the Jastrow figure received a new twist when Macpherson (2006) claimed that in Jastrow-type ambiguity, the nonconceptual content of visual experience also changes. (A distinct but substantially similar position is Siegel's 2010 generalized claim that conceptual content influences perceptual content.) Irrespective of whether those claims are true and what kind of penetrability they amount to, the claim that in figure 3 the switch is *merely* conceptual cannot even get off the ground. It is the nonconceptual representation of shape, in particular the 2½D sketch itself, that changes. For example, points A and B go from being at different distances on the z axis of Peacocke's nonconceptual scenario content to being equidistant on that axis. There

is no comparable modification of nonconceptual content in the the Jastrow figure's ambiguity.

A related strategy for countering cognitive penetrability claims is to use low-level perceptual illusions like the Muller-Lyer illusion (Fodor 1988; Pylyshyn 1999, 2003) to argue that the modules that yield visual experiences as outputs cannot be influenced by belief. But the Muller-Lyer illusion, like the chessboard illusion, are not *visually* revisable. In the chessboard illusion, the visual processes that interpret changes in light intensity in order to preserve light constancy turn out to be impenetrable. We cannot as it were 'reach down' from personal-level awareness into the brain processes that interpret light intensity data, tamper with them, and produce an experience of a different colour. However, we do seem to be able to do something like this with certain processes that yield volumetric representations, since for example when we switch from 2½D to 2D contents in figure 3, the orientation of line AB changes. When we do this we bring our perceptions into line with our beliefs—exactly what Fodor and Pylyshyn seek to deny by their use of lower-level illusions. Thus, to the extent that bottom-up picture perceptions are visually revisable at personal-level of awareness—with the exception of trompe l'oeils, which I will argue constitute rule-proving exceptions (§6)—such picture perceptions do not constitute lower level illusions.

Perhaps certain depiction theorists can come to the rescue of cognitive impenetrability. Some accounts of depiction hold that we have experiences of pictures as 2D and 2½D *simultaneously* (Wollheim 1987, 46; Gibson 1986, 282; Rock 2001, 98), which contradicts



the account given above of figure 3. In its extreme form—the claim that we *always* visually experience pictures as 2D and 2½D simultaneously—this proposal (which consists essentially of phenomenological claims) is in fact contradicted by the phenomenology: in bottom-up picture perceptions, which are the most frequent kind, the two experiences are separated by the considerable mental effort required to switch from one of the two visual experiences to the other. But in fact, even to say that we *sometimes* have a visual experience both of the surface and the picture-content seems wrong. In order to see a cube drawing as both flat and cubical at the same time, we would have to represent the intersection of the enclosed lines (point A) as occupying two locations at once. There are several ways of generating visual contents from pictures, but neither the natural, hard-wired ways nor the attention-driven visual interpretations make us represent a point at two different locations simultaneously. Although the visual system can yield separate and conflicting depth interpretations of scenes, each interpretation always seems to be *consistent* (Waltz 1975; Pylyshyn 1993, 99-107; Cutting and Massironi 1998 also give several illustrations of this visual assumption in their discussion of line interpretation). When vision cannot solve the problem of depth relations in a scene consistently, the result is an incomplete content and visual experience of only *part* of the picture at a time, as when we see the devil's pitchfork or certain drawings by Escher. Moreover, it is not part of the phenomenology of the picture-perception that we see the point in two places at once; on the contrary, it is very much part of the phenomenology that we see the point at one location *or* the other and that the two experiences are separated temporally by an 'attentional switch' (which itself seems to be positively experienced). In fact, in the case of

the simple outline drawing of a cube, there is little *else* to the phenomenology of the perception than the locating of these points in space.

In that case, what could explain the insistence of Wollheim, in particular, on the twofoldness thesis? Wollheim holds that the twofold experience of pictures is a single experience with two ‘aspects’. Perhaps we could give the following construal of that claim. Matthen (2005, 309-313) has argued that pictures do not engage the motion-guiding visual system, a visual system thought to be implemented in the dorsal pathway and to be distinct from the system dedicated to recognition and to representing object features (Ungerleider & Mishkin 1982; Milner & Goodale 1995). According to Matthen, the impact of this physical difference on the phenomenal character of picture perceptions is that they do not cause a ‘feeling of presence’ (2005, 306). So perhaps the dorsal system dedicated to navigation ‘knows’ that the picture is a more or less flat object, while at the same time the ventral system picks up the volumetric contents and depth relations from the picture’s surface. This would solve the apparently paradoxical nature of phenomenological accounts of twofoldness found not only in Wollheim (1987, 46) but also in Peacocke (1987, 386, 394) and in Gibson (1986, 282). Wollheim insists that the viewer ‘remains *visually aware* not only of what is represented but also of the surface qualities of the representation’ (1980, 216, italics added), even describing the experience as ‘twofold *attention*’. Perhaps this can be squared with the kind of twofoldness patched together in the last paragraph. Even so, a problem would remain: there is no room in this account of twofoldness for the fact that we consciously perform mental actions in order to switch from one set of visual contents to the other. If we really are visually aware of two contents at once, as Wollheim holds, they

cannot be the two sets of contents that we have trouble switching to and from when we perceive figure 3 (or, for that matter, figure 2).

Another attempt to explain Wollheim's convictions about twofoldness could be that they result from conflating two senses of 'seeing', a contentful and a purely causal sense: the surface causes the picture content and in this sense we see it, but when it does this, we don't see the surface *where* it is—since visually experiencing the content means experiencing different points of the surface at different distances. Once again, this does not account for Wollheim's particular position, since he insists that we do not see the surface in the bare sense that the surface is a causal stimulus, but that we visually experience it. Yet another explanation could be this, which concerns colour in particular. The colour that we place as a feature at a location, usually behind the picture-surface, is the colour caused by the surface. So one may hold that we have an accurate perception of the surface in respect of colour and at the same moment an inaccurate perception in respect of location. The *location*-component of the content is inaccurate because there is no blue *there*, while the *feature*-component is accurate because there is *blue* (but not where it is located). In that case, we do not see the colour in two places at once: we see it behind the picture-surface. This account would be compatible with typical phenomenological descriptions of perceiving pictures in Wollheim, as when he says that we 'marvel endlessly at the way in which line or brush stroke or expanse of colour is exploited to render effects or establish analogies that can only be identified representationally' (Wollheim 1980, 216). However, the solution cannot be extended to *shapes*. Colours (including non-chromatic ones in etchings and drawings) yield shape information, but adopting the solution for shapes

would again lead to the problem encountered earlier, of seeing the same point in two places at once. Thus, it's very tempting to conclude that the position defended by Gombrich and Cutting is right: we cannot perceive both the surface and the content of pictures at the same time.

The upshot of this discussion of simultaneity and twofoldness (theses to the effect that we have visual experiences of pictures as 2D and 2½D simultaneously) is that they do not offer a plausible alternative to the account given above of visual ambiguity in bottom-up, naturalistic picture perceptions. Such pictures can admit a form of visual ambiguity which implies that certain visual processes which lead from the 2D sketch to the 2½D sketch are penetrable by consciously driven processes. This modulation of visual processes does not just change *which* data is processed (as occurs in cases of attention-shifting) but *the way* visual data is processed. I think it unlikely that all bottom-up picture perceptions are subject to this form of ambiguity, but it suffices that *many* of them are to establish the penetrability of the relevant visual processes.

The robustness of the picture contents caused by naturalistic pictures—the difficulty we have in generating visual experiences with 2D contents while looking at naturalistic pictures—favours depiction and is an enabling condition for pictorial representation. For picture-perception as we know it to occur, we have to be able to continue to exploit the depth representations formed by earlier cognitive stages even *after* we have reached the stage at which we can form doxastic states. Otherwise, the mental representation of what the picture represents would evaporate each time we stopped making an effort to direct

spatial attention in the right way. Imagine the opposite scenario. If the 2½D visual interpretation did not prevail naturally each time we looked at the picture, perceiving pictures would be a painfully slow and attention-consuming activity; we would only be able to understand one part of a picture at a time with considerable effort. In that case, seeing the opaque cube-drawing as a cube would be as hard to do as it is now to see the cube-drawing as a flat surface. Picture perception would require great mental concentration on one region of the picture at a time. Higher-level attention is a valuable and energy-consuming resource for the brain, and thus it is unlikely that depiction would be as widespread an activity as it actually is. Instead, the 2½D visual interpretations prevail, and depiction as we know it—especially the profusion of different individual styles in drawing and painting—has emerged by exploiting that robustness. At the same time, admitting cognitively driven picture perceptions alongside bottom-up ones (see the next section) can explain both why we can use found objects as fortuitous pictures, and how we understand picture styles that exploit or frustrate visual procedures in different ways, either for aesthetic effect or due to technical limitations.

#### **4. Cognitively driven picture perceptions**

In the second form of visual ambiguity supported by pictures, the picture surface causes a veridical visual experience with 2D content in a bottom-up way by the time we reach phenomenal awareness; but it can subsequently support a visual experience of a 2½D scene if we agentively control spatial attention, optionally with backup from conceptually encoded information. The less ambiguously and more fluently a surface causes a 2D visual

representation, the greater the conscious effort required to use the surface to have an experience with volumetric and depth content. For example, in figure 4a, the visual system detects no occlusion. It is unlikely to detect a surface discontinuity because the change in colour is not sharp or regular enough. In the absence of other depth cues, the object is likely to suggest to the visual system a roughly plane, irregularly stained surface. (With some priming, however, we can see the figure as a convexity lit from the left.) Figure 4b supports an experience of the lower and upper halves as foreground and background (see §1). If we have trouble performing this segregation, it helps to *think* of the figure as a field with stalks against a gray sky. If the same object is viewed upside-down (4c), the figure-ground effect goes away, possibly because of an implicit assumption that light comes from above (see Ramachandran 2004). But in that case, we can reinstate the figure-ground segregation and the nonconceptual depth representation if we are given conceptual cues; eg, if we're told to see the figure as a sandy stretch in the foreground with reeds or shrubs in the background.

*a*

*b*

*c*

*Figure 4*

Both the phenomenal character and the nonconceptual content of visual perception change when the figures are seen as either segregated figure and ground or as a convexity, instead of as a stained flat surface. According to theories that defend perceptual impenetrability (Pylyshyn 1999; Pylyshyn 2007, 72; Raftopoulos 2009, 134), the processes that construct the nonconceptual volumetric representations in Marr's 2½D sketch are part of early vision and thus claimed to be impenetrable. Perceptual representations of depth are part of Marr's 2½D sketch. (The 2½D sketch is no less three-dimensional than the 3D sketch; the difference is that the 2½D sketch is 'a viewer-centered representation of the depth and orientation of the visible surfaces', the 3D sketch a representation whose 'coordinate system is object centered'; Marr 1982, 330.) But because *conceptual* information can determine which nonconceptual depth representations we have of the figures, this activity of visualizing or seeing-in is not just a case of perceptual ambiguity which can be contained within lower-level processes. Instead, it fits Pylyshyn's definition of cognitive penetrability: the visual experience of depth is 'altered in a way that bears some logical relation to what the person knows' (Pylyshyn 1999, 343)—namely, to conceptual information or memories of the textures of specific kinds of scenes. Thus, the processes that yield volumetric representations as outputs turn out to be cognitively penetrable.

Note that for figure 4b, while we *can* have conceptual representations (a field of wheat against a gray sky), perhaps we can also have a nonconceptual depth representation without having to use any concepts, solely on the basis of the figure-ground segregation. However, in that case, the segregation would not have to be performed bottom-up; it is possible to perform it *consciously* by imagining that there is a distance depth-wise between

the top and bottom parts of the figure.<sup>6</sup> This would count as a cognitively driven (personal-level, though not conceptual) form of amodal completion. In Kanizsa and Gerbino's terminology, the depth perception in figure 4c—and perhaps 4b, depending on the subject's perceptual set—could be described as a case of *represented* completion as opposed to perceptual completion. As Kanizsa and Gerbino put it, such a construal of the figure does not give us

the impression of being faced with something 'objective,' independent of us, not influenced by our will or our cognitive set. Indeed, properties such as phenomenal givenness and independence from the observer characterize a perceptual datum and distinguish it from a datum that is merely thought (Kanizsa and Gerbino 1982, 174).

Briscoe (2011) also calls similar completions 'cognitive'; and elsewhere uses the term 'make perceive' to describe comparable top-down cognitive activities: 'one engages in make-perceive when one projects or 'superimposes' a mental image on a certain region of the visually perceived world' (Briscoe 2008, 482).

---

<sup>6</sup>If we perform the figure-ground segregation, do we have the same visual experience of figure 4b with and without the concept? Siegel 2006, 2011a, thinks we do not; Lyons (2011, 305) thinks that the conceptual phenomenology may be 'a late experiential effect, leaving the nonconceptual early perceptual states unaffected but influencing the nondoxastic seemings'.



Like the claim about penetrability made in the last section, this one too is of a different kind to the claims made in Macpherson 2006. Macpherson discusses cases of perceptual ambiguity in which she argues that the phenomenal character of perception changes but the nonconceptual content can remain the same, whereas the cases of seeing-in described here bring about changes in content along with phenomenal character. This is because the cases discussed in Macpherson 2006 essentially concern mental rotations of represented shapes without altering those shapes, whereas the cases described here concern alteration of the shapes themselves (from 2D to 2½D or vice versa). On the other hand, the form of penetrability described here may be compatible with part of Macpherson's 2012 account of the effect of mental imagery on vision.<sup>7</sup>

The need for top-down cognitive influence in picture perception is not limited to fortuitous pictures. Paintings and drawings in many styles require subjects to make an active personal-level effort to visually experience their contents. This can be as a result of technical limitations like absent or imperfect perspective (in Roman wall paintings for instance), or deliberate choice in caricature and in partly abstract styles (like expressionism, cubism, and so on). Many of Turner's paintings, or parts of the paintings (of which the above figures are photographs) are in this category. While it is possible to perceive depth in those paintings, Ruskin held that Turner's paintings teach the eye to be 'innocent' about content and focus on traits of the picture surface; for Ruskin, as Gombrich

---

<sup>7</sup>That is, I think Macpherson's (2012) thesis that colour perception is cognitively penetrable is contestable (see Zeimbekis 2013), but I agree with the general point that mental imagery can penetrate vision (in shape perception, at least).

put it, ‘we do not even see the third dimension, only patches of colour and textures’ (1960, 238).

## 5. Picture perceptions and perceptual beliefs

To claim that cognitive and agentic states can modulate the outputs of visual processes is potentially to admit circularity into the justificatory relation between perceptual experience and belief. As Siegel (2012) puts it, if Jill visually experiences Jack’s face as angry *because* she believes that Jack is angry, then visually experiencing Jack’s face as angry can no longer justify her belief that Jack is angry. Siegel’s question applies in the following way to the forms of visual ambiguity outlined. On one hand, the experiences generated when we see figure 3 as flat, and when we see figures 4a-c as 2½D, are in some sense *visual* experiences. At the same time, those experiences are caused jointly by us—by agentively driven activities jointly with bottom-up processes. The suggestion is that we can contribute to deciding which visual experiences we have, and this would vitiate the ability of visual experience to justify belief.

Note that, in one sense, it is relatively trivial that we can decide which visual experiences to have. If we accept that we can never experience simultaneously *all* of the nonconceptual visual contents that a visual scene *could* cause, because we cannot focus attention on all parts of a visual scene at once,<sup>8</sup> then, given that we are capable of agentively directing the focus of visual attention, it follows trivially that we choose what to visually experience.

---

<sup>8</sup>For a description and plausible defense of this position see Naccache and Dehaene, 2007.

For example, if applying the concepts *pine* and *tree* to the same object required having *different* visual experiences of it, that would not require having *contradictory* contents and would not vitiate the justificatory relation: the visual experience which supports the concept *pine* would also support the concept *tree*; the concepts name kinds standing in a determination relation. But that is not what happens in cases of visual ambiguity—at least not in the ones described here.<sup>9</sup> The ambiguity of figures 3 and 4 allows us to have contradictory visual contents, so the threat is that those visual contents can support contradictory attributions or beliefs.

There are two ways to block the inference from cognitive penetrability to the vitiation of perceptual justification. The forms of penetrability described afflict picture perception, so one response would be to claim that even if processes which generate volume perception in *picture* perception are cognitively penetrable, volume perception processes in *object* perception are not penetrable. Another response would be to show that of the two states caused in each case of visual ambiguity, at most one state has the roles of causing and justifying beliefs—in other words, that only one state counts as a perception functionally and epistemically. In fact, these two questions turn out to be very closely related. I'll start from the second question, and this will give us the means to settle the first question.

If the capacity of visual perception to justify beliefs is vitiated by either of the forms of visual ambiguity outlined, it is unlikely to be the second form, in which visual processes naturally yield a veridical perception of the picture's surface as flat and we subsequently

---

<sup>9</sup>It does happen in Siegel's example of perceptions of pine trees.

use the surface as a visual prop to visualize a 3D scene. It immediately seems suspect to suggest that a *cognitively driven* act of visualizing could compete as a state for *perception's* causal and justificatory roles. When a fortuitous picture, like the wall stains described by Leonardo, naturally causes a visual experience of a flat surface, that experience causes a perceptual belief with the same content. Visual experiences of depth are produced after the formation of that perceptual belief by agentively controlling the focus of spatial attention. They are something we *do* akin to a mental action, not something that happens to us as the outcome of stimulus-driven and hard-wired processes functioning autonomously. Those states do not feel like perceptions and there is no reason to think that they have the causal role of causing perceptual beliefs which could compete with the belief caused by the initial, veridical visual experience.

Matters are more complicated when it comes to penetration of visual processes which support *naturalistic* picture perceptions. Remember that on the present hypothesis, in naturalistic picture perceptions—much as in object perception—by the time we reach phenomenal awareness, we have already formed a mental representation of a 2½D scene. The processes which cause the 2½D representation turn out to be cognitively penetrable: with mental effort, we can succeed in having a veridical experience of the picture, or parts of it at a time, as a flat surface. Thus, it seems that a subject could choose to have a veridical or a non-veridical visual experience at will, and the experiences would have contradictory contents. But this time, we cannot explain away the falsidical experience as a consciously driven act of visualizing; in fact, it is the experience caused by an act of visualizing that is veridical. Therefore, there seems to be only one way to avoid the

conclusion that subjects can have experiences with contradictory contents, which can justify contradictory beliefs. It is to argue that the 2½D visual experiences do not cause beliefs; in fact, to argue that as states, they do not fulfil the functional role of *perceptions* and in that sense at least do not constitute perceptions. I will sketch an explanation of why I think this may be so. If I am right, then the fact that vision is cognitively penetrable does not—at least in such cases—imply any vitiating of perception's role in justifying and grounding beliefs.

Stimulus-driven picture perceptions take a free ride on the processes that subserve object perception, so they have much in common with object perceptions, but that does not mean that the two are *identical*. The key difference is that in picture perception the brain does not calculate volumes and depth relations by using either binocular disparity from stereopsis or parallax from movement. According to Marr (1982, Ch. 3), stereopsis is largely responsible for generating the 'viewer-centered representation of the depth and orientation of the visible surfaces' (330) in object perception—the 2½D sketch which, as we saw, is likely to provide the nonconceptual content of visual experience. Picture perception, on the other hand, can exploit neither stereopsis nor parallax, and relies exclusively on the kinds of monocular depth cues described in §2. Therefore, binocular picture perception generates 2½D brain representations, and the nonconceptual contents of phenomenal awareness, not just by using *monoptic* perception, but by using a *subset* of the processes of monoptic object perception: not only is stereopsis absent, as it is in monoptic vision, but parallax from movement, which *is* available to monoptic object perception, is

also absent from picture perception. As brain states, picture perceptions are very different to object perceptions, if only for these reasons (additional ones will be given shortly).

These differences between the two kinds of visual states seem to be reflected by differences in their phenomenology. As Briscoe (2008) points out, the case of Susan Barry (reported by Sacks 2010), who had monoptic vision and only experienced stereoptic vision for the first time at an advanced age, suggests that the visual experiences caused by stereoptic and monoptic vision can be distinguished on phenomenological grounds:

I noticed the edge of the open door to my office seemed to stick out toward me. Now, I always knew that the door was sticking out toward me when it was open because of the shape of the door, perspective and other monocular cues, but I had never seen it in depth. It made me do a double take and look at it with one eye and then the other in order to convince myself that it looked different. It was definitely out there. [...] While I was running this morning with the dog, I noticed that the bushes looked different. Every leaf seemed to stand out in its own little 3-D space. The leaves didn't just overlap with each other as I used to see them. I could see the SPACE between the leaves. The same is true for twigs on trees, pebbles on the road, stones in a stone wall. Everything has more texture. (Sacks 2010, 125; quoted by Briscoe 2008, 473)

Susan Barry's reports of the phenomenology of her visual experiences suggest that depth is represented much less vividly<sup>10</sup> in monoptic object perceptions than in stereoptic ones—

---

<sup>10</sup> Whether or not this implies a change in representational content; a point I discuss below.

despite the fact that here, parallax *is* available to the subject, unlike in picture perceptions. She also reported that before her vision was corrected, she predicted that she could imagine what stereoptic vision was like, but retracted the claim after she experienced stereopsis (Sacks 2010, 122), which suggests that stereopsis acquainted her with a new kind of experience. Briscoe describes the difference between seeing monoptically and seeing stereoptically as a ‘dramatic influence of binocular depth information on the spatial phenomenal character of our visual experience’ (2008, 473). His explanation of the phenomenological difference is that after the brain calculates three-dimensional shape on the basis of binocular disparities in the light information reaching the retina, the information and processes are ‘not lost in our conscious visual experience of the object. Indeed [...] we can literally see the difference made by their presence (and absence) in the light available to the eyes’ (Briscoe 2008, 473).

How does this case throw light on picture perceptions? The difference between object perceptions and natural, stimulus-driven picture perceptions is that between (a) constructing the depth representations available to phenomenal awareness by means of stereopsis, parallax, and hard-wired processes that interpret monocular cues, and (b) constructing those representations only by means of monocular cues. Susan-Barry-type experiences are subserved by an intermediate state, (c): parallax plus monocular cues, but no stereopsis. Since there is a phenomenological difference between (a) and (c) that a subject whose physiology is working in the relevant respect (unlike Susan Barry before her vision was corrected) is able to distinguish, there is all the more reason to think that a

---

subject should be able to tell apart (a) normal object perception from (b) picture perception.

Apart from the suppression of stereopsis and parallax, there are other grounds for distinguishing (natural, stimulus-driven) picture perceptions from object perceptions, both as states and phenomenologically. As we saw in §3, Matthen (2005, 309-313) has argued that pictures do not engage the motion-guiding visual system, but only the visual pathway dedicated to recognition and to representing object features (the distinction is made by Ungerleider and Mishkin 1982, and Milner and Goodale 1995). Matthen claims that this physical difference is reflected phenomenologically by the fact that picture perceptions do not cause a 'feeling of presence' (2005, 306). Lack of engagement of the motion-guiding system is a functional effect. The absence of the brain states and processes that calculate disparity and parallax, the lack of engagement of the dorsal visual system, and increased reliance on monocular cues, may not affect only the phenomenology of picture and object perceptions: it may also mean that the causal antecedents for perception to play its functional role of causing perceptual beliefs are not satisfied, something which could directly prevent even natural, stimulus-driven, picture perceptions from causing perceptual beliefs. Another feature of picture perception which is compatible with both the lack of stereopsis and parallax, and the lack of engagement of the motion-guiding system, could be the lack of subject independence. Subject independence is claimed by Siegel (2006b) to be part of the phenomenal character of object perceptions. This feature is absent from picture perceptions not only because the locations of objects represented in picture perception are insensitive to parallax, but also because, as Cutting has shown, shapes



represented in picture contents withstand considerable changes in the viewing angle (Cutting 1987).

I have given a number of reasons why the visual *states* caused by natural, stimulus-driven picture perceptions should differ from object perceptions, and some evidence and reasons why the *phenomenology* of the states should also differ. Now, is this difference between object perceptions and natural picture perceptions a difference of sensory or cognitive phenomenology?

According to Sacks (2010, Ch. 5), the suppression of stereoptic representations of depth can have an impact on representational content itself; for example, it can lead to miscalculations about the shape and location of objects. And Briscoe, as we saw, describes the phenomenological difference between monoptic and stereoptic vision as being due to the ‘influence of binocular depth information on the spatial *phenomenal character* of our visual experience’, and writes that ‘we can *literally see* the difference’ made by calculation of shape from binocular disparities (2008, 473; my emphasis). This suggests that the phenomenological contrast between stereoptic and monoptic vision is due to differences of representational content. Nevertheless, differences of representational content may not exhaust the phenomenological differences between stereoptic and monoptic vision, nor those between object perceptions and natural picture perceptions. One reason for this is provided by Matthen’s proposal. Lack of engagement of the motion-guiding system by pictures would not have effects on the object features represented visually, but it would have the effect of suppressing the epistemic feeling of presence, according to Matthen. In

that case, the contrast between the phenomenologies would not co-vary with changes to representational content. That does not necessarily mean that the missing ingredient has to be a form of cognitive phenomenology.<sup>11</sup> Perhaps the epistemic feeling of presence supervenes on early, unconscious visual processes; that is certainly suggested by Matthen's (2005) account (see §3).<sup>12</sup> A similar point applies to subject-independence, which Siegel considers part of perceptual experience: if it is absent from picture perceptions, that does not mean that picture perceptions and object perceptions differ in terms of cognitive phenomenology; yet nor is the difference one of visual contents.

When reporting her experiences, Susan Barry contrasts representing the orientation of a door 'because of the shape [and] perspective' to 'seeing it in depth', and representing the 'overlap' of leaves to 'seeing the space between the leaves'. There are different ways to understand these passages. It is true that the closer objects are to the viewer, the more binocular disparity can reveal of the space that would be occluded by either one of the monocular views. On one reading, this is what Susan Barry is reporting, and it is a change in phenomenology that co-varies with a change in visual representational content. However, such changes to representational content can also be obtained through parallax, so their phenomenology should have been familiar to Susan Barry before her vision was corrected—they would not constitute a novel form of phenomenological vividness or

---

<sup>11</sup> Thanks to an anonymous O.U.P. reviewer for asking me to pursue this question.

<sup>12</sup> See also Dokic and Martin, this volume, who describe epistemic feelings as 'the output of a monitoring process which involves implicit inferences from a set of internal cues, such as availability of partial information or fluency'.

salience. So it is possible that the passages report a phenomenological change that does not co-vary with representational content. Briefly, Susan Barry's reports may not be about seeing more occluded space than before, but about seeing space differently. To conclude, perhaps stereoptic and monoptic vision differ phenomenologically not only because they have different visual contents, but also in ways that overflow any differences between their visual contents.

## **6. Illusory picture perceptions**

This account of the causal and epistemic role of picture perceptions is supported by a consideration of pictures that *do* cause illusions: trompe l'oeils. 'Trompe l'oeil' is an ambiguous term which sometimes designates a picture and sometimes the effect of a picture. Here, I use the term for pictures in contexts in which they succeed in causing higher-level illusions, typically signaled by a sense of surprise when the illusion is dispelled. (I do not use the term for paintings that are extremely naturalistic but appear in contexts or conditions in which they do not produce the illusion, like William Harnett's *Old Models*; see Goldstein 2005, 354.) Figure 5, while it is only a photograph of a picture (a mural), provides a good illustration.<sup>13</sup>

---

<sup>13</sup>Thanks to John Pugh for generously granting me permission to reproduce this work.

*Figure 5*

Trompe-l'oeils, for as long as they work, make us ascribe properties represented by the picture to locations in the space represented by the picture. The picture contents—nonconceptually represented shapes, colours and textures, concepts for properties and kinds, and object files for objects to bind the features and instantiate the properties—reach doxastic-level awareness wholesale, and form the structured contents of a perceptual belief. We could form such a belief if we turned a corner and came upon the building in figure 5. That in turn implies that the phenomenal character of the experience, and the causal role of the state, are indiscernible from those of an object perception.

Trompe-l'oeils do not get that far under any contextual conditions. A precondition for their success is that the picture should not be presented as a picture, something which would affect 'perceptual set': the preparedness for certain categories of object, scene or action, which is capable of biasing the visual processing of depth relations such as figure-ground segregation and could therefore defeat the trompe l'oeil's ability to cause an illusory belief (see Vecera 2000, 367-370, for an overview of the concept of perceptual set, and Peterson and Hochberg, 1983, for the claim that perceptual set can affect figure-ground segregation). This is why successful trompe l'oeils usually appear in particular visual

contexts, which could be described as unexpected in one sense, but also—in another sense—as consistent with expectations. On one hand, they benefit from being embedded in contexts where we do *not* expect to see a picture, which is why they are often painted on architectural features like external or internal walls, ceilings and columns. On the other hand, the *content* of the trompe l'oeil should be consistent with the context in which it is embedded, which is why contents frequently consist of fluting, panels, mouldings, and other features consistent with the architectural setting. When such contextual conditions are met, even coarsely made pictures which are not naturalistic can momentarily fool the mind into confusing the picture perception with an object perception.

But the condition which is of direct interest here is that trompe l'oeils work as long as the absence of binocular disparity can go undetected and as long as parallax is neutralized, if we are not stationary. Thus, the ability of the picture to cause the illusionistic effect increases with the viewer's distance from it, and the distance required is itself proportionate to the depth represented in the picture contents: the shallower the scene represented, the smaller the binocular disparities it would cause and the harder it is to detect their absence. A frequently used theme in trompe l'oeils are false window casings painted high on a building, which meet all of these conditions, as well as the earlier conditions connected with perceptual set. When trompe l'oeil illusions are dispelled, it is usually because the absence of parallax is detected as we move relative to the wall, column or ceiling in which the picture is embedded.

Once the illusion is dispelled, the experience of the picture seems to change. For example, when we realize that the lines suggesting the presence of a window casing are planar and that there are no window casings where we represent them, we can still mentally represent the casings—in other words, the *content* of the picture—as we look at the building. Support for this claim comes from the phenomenon that Kubovy (1986) calls ‘robustness of perspective’: picture perception is relatively insensitive to changes in the angle from which the picture is viewed; according to Cutting (1987), pictures can be rotated up to 22 degrees (on their vertical axis) without this deforming the three-dimensional representations they cause. This is a very substantial degree of rotation, so we should be able to represent picture contents throughout the movements that make us detect the lack of parallax. What changes throughout the movements is the *illusory* nature of the picture contents, not the contents themselves. The fact that picture perceptions and trompe l’oeil illusions represent the same contents under different psychological modes or attitudes corroborates the claim that picture perceptions are distinct from perceptual illusions.

Therefore, trompe l’oeils are rule-proving exceptions to the theory that picture-perceptions and object-perceptions have different phenomenal characters. They are *exceptions* because they produce picture perceptions which are momentarily indiscernible from object perceptions. But they are *rule-proving* exceptions because what makes them indiscernible from object perceptions is the fact that we do not detect the absence of binocular disparity and parallax. As such, they confirm the thesis that lack of stereopsis and parallax underlie the phenomenal difference between the remaining cases of bottom-up picture perception, and object perceptions.

## 7. Conclusion

A key part of vision—the processes that generate volumetric shapes and depth relations from monocular cues—is cognitively penetrable, with the result that what subjects visually experience can be a function of conscious acts of visualizing and semantic information. One would expect the cognitive penetrability of those processes to damage their capacity to provide warrant and justification for beliefs. However, the epistemic situations in which subjects find themselves provide no scope for cognitive penetrability to cause such damage. Consider a simple illustration of these epistemic situations. When we view Rubin's vase-face figure, the ambiguity is restricted to visual experiences; it does not translate into any ambiguity at the level of perceptual beliefs because neither of the visual experiences causes a perceptual belief that there is either a vase or a pair of faces located before us. Issues of warrant and justification simply do not emerge. They *would* emerge only for visual experiences of the figure as a flat, stained surface—and in that case, in a way that does not damage vision's capacity to provide perceptual warrant. The same applies to all the ambiguous figures used in the penetrability literature. To the extent that they are pictures, they can support forms of visual ambiguity that imply cognitive penetrability; but not in ways that entail the epistemological consequences usually expected of cognitive penetrability.

Could we conclude that the processes which generate volume perception are cognitively penetrable *only* during picture perception, not in the wider context of the processes that take place in object perceptions?

Think of object perceptions on a range. At one end are cognitively driven picture perceptions; next are pictures with weak depth cues; after that, naturalistic pictures with good depth cues, followed by trompe l'oeils; and finally, object perceptions in good viewing conditions. Object perceptions would be even harder to visually experience as 2D than trompe l'oeils, which are already very hard to visually experience as 2D. Stereopsis, visual exploration using parallax, and possibly the engagement of the dorsal stream's motion-guiding systems, would make it so difficult to have such visual experiences that we would have to actively visualize the scene before us as being 2D with a great mental effort. If it was possible to have such a state, it would have to be by keeping still to prevent the effect of parallax and viewing the object through one eye to stop the brain calculating binocular disparity; and even then, we would still have monocular cues to contend with, so such a state could still not emerge solely on the basis of stimulus-driven and hard-wired processes.

Now, if we could *not* get to the point of having such a state, the processes which generate volume perception would not be cognitively penetrable in visual *object* perception taken as a wider set of visual processes than picture perception; they would only be penetrable in picture perception. This would be an interesting outcome, because challenges to impenetrability and modularity from visual ambiguity always use pictures as examples:



pictures are used to show that there is penetration of some process, and then the conclusion is that that process is generally penetrable. The conclusion would be false if in object perception we simply could not get ourselves into visual states whose contents were two-dimensional.

But suppose that we *could* have such states. It is plausible that we can, because viewing conditions can be far from optimal in object perception (especially for distant objects, for the reasons seen in the discussion of trompe l'oeils), and the visual system is made to be able to deal with such conditions (just as it seems to do when we can make sense of non-naturalistic pictures). Then—for the same reason that cognitively driven *picture* perceptions do not cause perceptual beliefs—these consciously sustained mental representations of *real scenes as flat* would not cause perceptual beliefs either. They would be visual experiences that we ourselves actively sustain as agents by performing acts of visualizing, not states that happen to us like purely stimulus-driven and hard-wired visual contents are. Yet, it would remain that the processes which generate volume perception in object perception are cognitively penetrable; it is just that the states resulting from penetration would not cause beliefs, and could therefore not vitiate the justification of beliefs by visual perception.

In either case—whether the visual processes that construct volume and depth are penetrable only in picture perception, or also in object perception—it transpires that we can admit the cognitive penetrability of fundamental visual processes without threatening

the epistemic relation between visual perception and belief. So we do not have to deny cognitive penetrability to uphold the perceptual justification of belief.

## References

Abell, C. and Currie, G. (1999). Internal and External Pictures. *Philosophical Psychology* 12.4: 440-441.

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychological Review*, 94, 115-147.

Biederman, I. (1995). Visual object recognition. Kosslyn, S. and D. N. Osherson (Eds.), *An invitation to cognitive science*. MIT Press. 121-165.

Biederman, I. and Kim, J. (2008). 17000 years of depicting the junction of two smooth shapes. *Perception* 37: 161-164.

Block, N. (1990). Consciousness and accessibility. *Behavioral and Brain Sciences* 13, 596-598.

Block, N. (1997). On a confusion about a function of consciousness. In Block, N., Flanagan, O. and Güzeldere, G. (Eds.) *The Nature of Consciousness: Philosophical Debates*. MIT Press.

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences* 9. 2, 46-53.

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

- Briscoe, R. (2011). Mental imagery and the varieties of amodal perception. *Pacific Philosophical Quarterly* 92.2, 153-173.
- Bulthoff, H. and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America* 89: 60-64.
- Campbell, J. (2002). *Reference and Consciousness*. Oxford University Press.
- Cavanagh, P. (2005). The artist as neuroscientist. *Nature* 434.17: 301-307.
- Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science* 55: 167-187.
- Cutting, J. (1987). Rigidity in cinema seen from the front row, side aisle. *Journal of Experimental Psychology* 13.3, 323-334.
- Cutting, J. and Massironi, M. (1998). Pictures and their special status in cognitive inquiry. In Hochberg, *Perception and Cognition at Century's End*. Academic Press.
- Delk, J. and Fillenbaum, S. (1965). Differences in perceived color as a function of characteristic color. *American Journal of Psychology* 78.2, 290-293.
- Edelman, S., Bulthoff, H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research* 32.12, pp. 2385-2400.
- Fodor, J. (1983). *The Modularity of Mind*. MIT Press.
- Fodor, J. (1988). A Reply to Churchland's 'Perceptual Plasticity and Theoretical Neutrality'. *Philosophy of Science* 55.2: 188-198.

Gibson, J. (1986). *The Ecological Approach to Visual Perception*. New York, Taylor and Francis.

Goldstein, B. (2005). Pictorial perception and art. In Bruce Goldstein (Ed.), *Blackwell Handbook of Sensation and Perception*, 344-378.

Gombrich, E. [1960] (1984). *Art and Illusion: a Study in the Psychology of Pictorial Representation*. London: Phaidon Press.

Kanizsa, G., and Gerbino, W. (1982). Amodal completion: Seeing or thinking? In B. Beck (ed.), *Organization and Representation in Perception*, 167-190. Hillsdale, N.J.: Erlbaum.

Kosslyn S. (1994). *Image and Brain: The Resolution of the Imagery Debate*. MIT Press.

Kourtzi, Z., Grodd, W., and Bulthoff, H. (2003). Representation of the Perceived 3-D Object Shape in the Human Lateral Occipital Complex. *Cerebral Cortex* 13: 911-920.

Kubovy, M. (1986). *The psychology of perspective and Renaissance art*. London, Cambridge University Press.

Lamme, V. (2000). Neural Mechanisms of Visual Awareness: A Linking Proposition. *Brain and Mind* 1: 385-406.

Lamme, V. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences* 7.1, 12-18.

Leonardo da Vinci. (1989). *On Painting*. Yale University Press.

Levinson, J. (1998). Wollheim on Pictorial Representation. *Journal of Aesthetics and Art Criticism* 56.3: 227-233

- Logothetis, N., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology* 5 (5), 552-563.
- Lopes, D. (1996). *Understanding Pictures*. Oxford University Press.
- Lyons, J. (2011). Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues* 21: 289-311.
- Macpherson, F. (2006). Ambiguous Figures and the Content of Experience. *Noûs* 40.1: 82-117.
- Macpherson, F. (2012). Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84(1): 24-62.
- Marr, D. [1982] (2010). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press.
- Matthen, M. (2005). *Seeing, Doing and Knowing. A Philosophical Theory of Sense Perception*. Clarendon Press, Oxford.
- Milner, A. and Goodale, M. (1995). *The Visual Brain in Action*. Oxford University Press.
- Naccache, L. and Dehaene, S. (2007). Reportability and illusions of phenomenality in the light of the global neuronal workspace model. *Behavioral and Brain Sciences* 30: 518-520.
- Peacocke, C. (1987). Depiction. *Philosophical Review* XCVI: 383-410.
- Peacocke, C. (1992). *A Study of Concepts*. MIT Press.
- Peterson, M. and Hochberg, J. (1983). Opposed-set measurement procedure: A quantitative analysis of the role of local cues and intention in form perception. *Journal of Experimental Psychology: Human Perception and Performance* 9: 183-193.

- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences* 22: 341-423.
- Pylyshyn, Z. (2003). *Seeing and Visualizing: It's Not What You Think*. MIT Press.
- Pylyshyn, Z. (2007). *Things and Places: How the Mind Connects with the World*. MIT Press.
- Raftopoulos, A. (2009). *Cognition and Perception. How do Psychology and Neural Science Inform Philosophy?* MIT Press.
- Raftopoulos, A. (2011). Ambiguous figures and representationalism. *Synthese* 181.3: 489-514.
- Ramachandran, V. and Rogers-Ramachandran, D. (2004). Seeing Is Believing. *Scientific American - Mind*, January 2004: 100-101.
- Sacks, O. (2010). *The Mind's Eye*. New York: Knopf.
- Schier, F. (1986). *Deeper into Pictures: an Essay on Pictorial Representation*. Cambridge University Press.
- Siegel, S. (2006a). Which properties are represented in perception? In Gendler and Hawthorne, eds., *Perceptual Experience*. Oxford: Oxford University Press.
- Siegel, S. (2006b). Subject and object in the contents of visual experience. *Philosophical Review* 115(3): 355-388
- Siegel, S. (2010). *The Contents of Visual Experience*. New York: Oxford University Press.
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs* 46(2):201-222.

Tarr, M. (1995). Rotating objects to recognize them: a case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review* 2.1: 55-82.

Tarr, M. and Bulthoff, H. (1998). Image-based object recognition in man, monkey and machine. *Cognition* 67, 1-20.

Tarr, M.J. and Pinker, S., (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology* 21.28: 233-282.

Tye, M. (1991). *The Imagery Debate*. MIT Press.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition* 67, 21-44.

Ungerleider, L. and Mishkin, M. (1982). Two Cortical Visual Systems. In Ingle, D., Goodale, M. and Mansfield, R. (Eds.), *Analysis of Visual Behaviour*. MIT Press. 549-586.

Vecera, S. (2000). Toward a Biased Competition Account of Object-Based Segregation and Attention. *Brain and Mind* 1: 353-384.

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

Walton, K. (2002). Depiction, Perception, and Imagination: Responses to Richard Wollheim. *Journal of Aesthetics and Art Criticism* 60.1: 27-35.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston (ed.), *The Psychology of Computer Vision*, 19-91. New York, McGraw-Hill.

Wollheim, R. (1980). *Art and Its Objects*. Cambridge University Press.

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

Wollheim, R. (1998). On Pictorial Representation. *Journal of Aesthetics and Art Criticism* 56.3: 217-226.

Zeimbekis, J. (2012). *Pictures, Perception and Meaning*. Manuscript.

Zeimbekis, J. (2013). Color and cognitive penetrability. *Philosophical Studies*, 165 (1) : 167-175.