

*Vision, Action, and Make-Perceive**

ROBERT EAMON BRISCOE

Abstract: In this paper, I critically assess the enactive account of visual perception recently defended by Alva Noë (2004). I argue *inter alia* that the enactive account falsely identifies an object's apparent shape with its 2D perspectival shape; that it mistakenly assimilates visual shape perception and volumetric object recognition; and that it seriously misrepresents the constitutive role of bodily action in visual awareness. I argue further that noticing an object's perspectival shape involves a hybrid experience combining both perceptual and imaginative elements – an act of what I call 'make-perceive.'

1. Introduction: The Sense-Datum Theory of Perception Redux

In his recent *Action in Perception* (2004), Alva Noë argues that influential criticisms of the sense-datum theory by Austin 1962, Gibson 1979, Strawson 1979, and others actually 'leave unscathed what is really the sense-datum theory's core idea: that perceiving is a way of finding out how things are from how they look or sound or, more generally, appear.... Looks, sounds, feels — appearances generally — are perceptually basic' (2004, 81). According to Noë, an object's visually apparent shape is the shape of the patch that would occlude the object on a plane perpendicular to the line of sight, i.e., the shape of the patch projected by the object on the frontal plane in accordance with the laws of linear perspective. Noë calls this the object's 'perspectival shape' (P-shape). An object's visually apparent size, in turn, is the size of the patch projected by the object on the frontal plane. Noë calls this the object's 'perspectival size' (P-size).¹ Appearances are perceptually basic because in order to see an object's actual spatial properties it is necessary both to see its P-properties and to understand how its P-properties would vary (undergo transformation) with variation in one's point of view:

To see a circular plate from an angle, for example, is to see something with an elliptical P-shape and it is to understand how that perspectival shape would vary as a

Address for correspondence: Robert Eamon Briscoe, Philosophy Department, Loyola University, 6363 Saint Charles Avenue, New Orleans, LA 70118, USA.

Email: rbriscoe@gmail.com

* I wish to acknowledge my indebtedness to Avner Baz, Zeynep Çelik, Michael Cohen, Daniel Dahlstrom, Daniel Dennett, Juliet Floyd, Aaron Garrett, Larry Hardesty, Ruth Millikan, Lisa Mosier, Alva Noë, and Jon Winawer for helpful discussions of the enactive account and/or various topics in vision science; to Roland Fleming and Pierre Jacob, who provided detailed comments and suggestions that lead to many significant improvements; and to James Conant and John Haugeland for arranging a visiting scholar position at the University of Chicago in fall 2005 that enabled me to write an early version of this paper.

¹ Note that there are infinitely many planes, each at different distance in depth, on which an object could be occluded by a suitably shaped patch. It follows that P-size is relative to plane of occlusion. (Compare the sizes of the patches that would occlude the Eiffel Tower on planes 50 and 5,000 feet from its base.) Noë is aware of this difficulty. Nonetheless he maintains that 'there is a single apparent size of an object – namely, the unique way that an object looks with respect to size from a particular position. This is secured by phenomenology' (2004, 84). The problem is that the notion of P-size was presumably intended as a theoretical replacement for the notion of 'apparent size' or 'size in the visual field.' So, if phenomenology does in fact pre-theoretically secure the claim that an object has but a single apparent size, then this just suggests that P-size is a bad theoretical replacement for apparent size.

function of one's (possible or actual) movements with respect to the perceived object (2004, 84).

Similarly:

To see the actual size of a thing is to see how its perspectival size varies as we move (2004, 84).

Seeing, as thus characterized on Noë's enactive, sensorimotor account of visual perception, is a 'two-step' process: 'How they (merely) appear to be plus sensorimotor knowledge gives you things as they are' (2004, 164).

The enactive account, it seems clear, bears more than a superficial resemblance to the sense-datum theory of perception. Crucially, both the enactive account and the sense-datum theory adhere to the following *phenomenological* and *perceptual* claims:

Phenomenological: An object's visually apparent shape is the shape of the patch that would occlude the object on a plane perpendicular to the line of sight. 'All bodies, which discover themselves to the eye,' as David Hume puts it, 'appear as if painted on a plain surface' (1740/1975, 1.2.5). Thus, when viewed from an oblique angle, round coins are supposed to appear elliptical, rectangular tabletops are supposed to appear trapezoidal, and so on. It follows from the phenomenological claim that visually apparent shapes are uniformly 'flat' or 2D.

Perceptual: Whenever an object visually appears Σ -shaped to a perceiver, then this is because the perceiver *sees* something – an 'appearance' or 'look' – that really is Σ -shaped. Thus Noë writes: 'P-properties are themselves *objects of sight*, that is, things that we see. They are visible. From where you stand, you can see the P-shape of the plate, and you can distinguish this from its actual shape' (2004, 83). Note that the perceptual claim is not merely that we see or are otherwise aware *that* a certain 2D ellipse would occlude the plate when viewed from an angle. The latter claim would not imply that the plate itself actually looks elliptical to us.

In the writings of the sense-datum theorists, the two claims serve to motivate what A.D. Smith has called 'One of the most notable features of both philosophy and psychology throughout the eighteenth and nineteenth centuries,' namely, 'the almost universal denial that we are immediately aware through sight of objects arrayed in three-dimensional space' (2000, 481). On this view, the immediate objects of visual perception are, strictly speaking, rather 2D, colored expanses. 'We see nothing but flat colours,' as John Ruskin puts it in a well-known passage from *The Elements of Drawing* (1857/1971, 27).

In Noë's recent 'Real Presence' (*forthcoming*), the two claims serve to motivate the view that an object's 3D shape and voluminousness are 'present as absent' in visual experience. Consider the case of viewing a silver dollar slanted in depth. According to Noë, the coin's circular, 3D shape is *absent* from our visual experience – 'there is,' he says, 'no seeing the coin's circularity from an angle' (ms. 17) – because the coin's visually apparent shape is its elliptical, 2D P-shape. At the same time, the coin's 'strictly unperceived circularity' (ms. 7) is *present* or, as Noë sometimes says, 'available' in our visual experience because we are able to see the coin's P-shape and we understand how its P-shape would vary as we move. In general, such movement-dependent variation in an object's 2D P-shape and our implicit knowledge thereof constitutes the 'real presence' of the object's 3D shape in visual experience. Or so Noë

argues.

In this paper, I critically examine and assess the enactive account of visual perception. In §2, I begin by criticizing Noë's proposal that in order to see an object's 3D shape it is necessary to possess knowledge of the object's 'sensorimotor profile,' of the way the object's 2D P-shape would undergo change with variation in one's point of view. This proposal, I argue, mistakenly assimilates visual shape perception, i.e., our ability to *see* the orientation of an object's visible surfaces in depth, to volumetric object recognition, i.e., our ability to *identify* the object's volumetric shape from any of indefinitely many different possible perspectives on the object. This error, I then try to show, is in fact symptomatic of a broader confusion about the role of bodily action in spatially contentful visual perception.

In §3, I argue that there are good empirical reasons to reject the claim – strikingly at variance with the post-Gibsonian tradition in vision science – that an object's visually apparent shape is identical with its 2D P-shape. Contrary to pronouncements made by Noë, I show that there is no credible evidence that 'genuinely naïve' perceivers would in fact take round coins presented at an angle to be elliptical. Indeed, studies of neonatal vision suggest that just the opposite is the case. In §4, I then criticize the weaker phenomenological claim that, in seeing 3D objects, we are *able* to see their P-shapes. In general, P-shapes are not properties that 3D objects (as opposed to pictures of those objects) can literally be seen to instance. That said, I allow that there is a sense in which we are sometimes able to notice an object's P-shape. Purposing to notice an object's P-shape, I speculatively propose, involves an act of what I call 'make-perceive.' To make-perceive an object's P-shape is to project or superimpose an image of the object's 2D P-shape constructed 'off-line' in imagination on one's 'on-line' visual experience of the object's 3D shape. Noë, I suggest, mistakenly identifies the shape of a 2D patch that one *imagines* while viewing the object with the shape that the object itself visually appears to instance.

According to Noë, P-properties are 'perceptually basic' (2004, 81). The enactive account, however, takes for granted our ability perceptually to individuate or 'segment' an object's P-shape, i.e., its perspective projection on the frontal plane. In §§5-6, I show that phenomenological and experimental studies provide compelling evidence that our first, conscious visual awareness of the world is perceptually organized in terms of visible surfaces arrayed and oriented *in depth*. Indeed, our ability perceptually to individuate discrete P-shapes on the frontal plane, I argue, is psychologically dependent on our ability to perceive the 3D organization of the scene projected on the plane. But, if this is right, then P-shapes cannot play the perceptually basic role conferred on them by the enactive account. In §7, I conclude with some brief remarks on the 'perspectival' character of visual experience.

2. Visual Shape Perception vs. Volumetric Object Recognition

I begin with the claim that in order to see an object's 3D shape it is necessary to possess knowledge of the object's 'sensorimotor profile,' of the way the object's appearance, i.e., its 2D P-shape, would undergo transformation with variation in one's point of view. Noë writes:

The sensorimotor profile of an object is the way its appearance changes as you move with respect to it (strictly speaking, it is the way sensory stimulation varies as you move). All solid, opaque objects have sensorimotor profiles in just this sense. As we get to more complicated forms, such as animal bodies, plants, and so forth, the mathematics needed to determine the sensorimotor profile of an object gets more complicated. Our visual perceptual skills, however, are that sophisticated, encompassing these complex (but ultimately manageable) relationships.

In this way the enactive approach explains the perceptual presence of shape, that

is, how visual experience can present objects to us as having three-dimensionality, volume, and shape (2004, 78).

The approach is illustrated by means of the following example:

When you move with respect to a plate, its profile changes. ...Our appreciation of its actual shape consists in our perception of its profile and our understanding of the way the profile, or apparent shape, depends on movement. We may say, in a case such as this, that we are able to experience the shape of the plate, to see it, because we grasp, implicitly, the sensorimotor profile of the plate. Our grasp of the plate's sensorimotor profile makes its shape available in experience (2004, 78).

Seeing an object's actual, 3D shape, on this view, then, crucially involves matching its 2D P-shape to a certain (presumably learned) sensorimotor profile. Whether or not one has matched the object's P-shape to the correct sensorimotor profile – and, so, whether or not one has formed the correct sensorimotor expectations – only becomes apparent as one moves in relation to the object.

Before proceeding, it is necessary to point out that assessment of the enactive account is significantly complicated by Noë's tendency to speak interchangeably of distal P-properties (or 'looks'), on the one hand, and physicalistically characterizable, proximal sensory stimulations, on the other. Thus, in the passage quoted above, he writes: 'The sensorimotor profile of an object is the way its *appearance* changes as you move with respect to it (strictly speaking, it is the way *sensory stimulation* varies as you move).' The difference in locution, however, is by no means inconsequential. Indeed, it isn't *prima facie* obvious how the different characterizations are to be related. P-properties, according to the enactive account, are distal, relational properties of the objects we see. 'If there is a mind/world divide...', Noë writes, 'then P-properties are on the world side of the divide' (2004, 83). Moreover, the enactive account clearly assumes that they are visible: 'P-properties are themselves *objects of sight*, that is, things that we see' (2004, 83). Sensory stimulations, by contrast, are proximal, subpersonal vehicles of visual perception. They are not objects of sight. Quite different, if not incommensurable, notions of 'sensorimotor profile' and, so, of 'sensorimotor understanding' would thus seem to be implied by the two characterizations.

Putting this preliminary interpretative difficulty to the side, the first, and most general, objection to the enactive account, I would suggest, is that it mistakenly assimilates the theoretical problems respectively posed by visual shape perception and volumetric object recognition. In consequence, the enactive account seriously 'over-intellectualizes' perception: it elides the cognitive requirements of *seeing* an object's visible 3D shape with the cognitive requirements of being able to *identify* the object's volumetric shape from any of indefinitely many different possible perspectives on the object.

Generally speaking, the problem of 3D visual shape perception is the problem of how – on the basis of multiple, independently variable sources of information in the light – we are able to see the orientation in depth of an object's visible, i.e. non-occluded, surfaces. Sources of depth-specific information include binocular disparity (stereopsis), motion parallax, perspective gradients, occlusion cues, texture, reflections, shadows, and, to a lesser extent, accommodation, convergence, and 'blur' (see §3 below). Explaining how perception of 3D shape is enabled by nonconscious processing of these sources of information is one of the fundamental problems of 'mid-level' vision or 'surface representation' as it is sometimes called (Nakayama *et al.* 1995, Nakayama 1999, Anderson 2003, Pylyshyn 2003).

In contrast, the problem of volumetric object recognition is the problem of how we are able to identify an object's volumetric (solid) shape on the basis of any of

indefinitely many possible 2D ‘images’ or ‘views’ of the object.² E.g., we are reliably able to identify an object as a coffeepot, or a car, or a chair, whether we see it from in front, from above, from the side, etc. Since this flexible identificatory ability is not wholly input-driven, i.e., is not wholly dependent on information in the light available to the eyes, but rather also depends on stored, object-specific knowledge, volumetric object recognition is standardly regarded as a ‘top-down’ process. Explaining how the process is psychologically implemented is one of the fundamental problems of high-level vision. For general discussion, see Marr 1982, Kosslyn 1994, and Ullman 1996.

It seems reasonably clear that our ability to perceive the orientation of an object’s visible surfaces in depth does not in general depend on our having the ability to identify its volumetric shape as that of a previously encountered object (or kind of object). Indeed, this is why we are able accurately to perceive the completely novel shape of a rock formation at the beach, or a cloud, or an abstract sculpture. Failure or inability to identify an object thus need not by itself imply any properly *visual* error. There is ample experimental evidence in support of this view. Numerous neuropsychological studies have shown that visual shape perception may be pathologically dissociated from object recognition in consequence of severe brain damage. Thus subjects with associative visual agnosia may have profound deficits in abilities to recognize objects while retaining normal visual depth perception and normal visuomotor abilities (Milner & Goodale 1995/2006, Farah 2001). This strongly suggests that stored knowledge about an object’s volumetric shape is not in general necessary in order to see the way the object’s visible surfaces are oriented in depth.³ (Which, I should note, makes preeminently good sense from a *phylogenetic* standpoint: it seems highly likely that relatively low-level, general-purpose strategies for recovering 3D surface layout from information in the light were phylogenetically in place prior to and, so, available to be exploited in the development of high-level abilities to recognize object-specific, volumetric shapes. For discussion of this point, see Millikan 2004, chap. 9).

The enactive account, however, clearly seems to assimilate the problem of how we see an object’s 3D shape to the problem of how we recognize its volumetric shape across different possible views of the object. This is most evident when, in chapter three of *Action in Perception*, Noë identifies an object’s sensorimotor profile with what Jan Koenderink and Andrea van Doorn call the object’s ‘visual potential’ or ‘aspect graph’ (Koenderink & van Doorn 1979, Koenderink 1984). In the object recognition literature, an aspect graph is a ‘viewer-centric’ representation of all topologically distinct, 2D views of an object in a network format (figure 1). To each view of the object, or aspect, there corresponds a node in the network. Connections between nodes represent translations between adjacent views of the object in 3D viewpoint space and thus mirror the way the object’s aspect changes as the result of movement. The task of visual object recognition, on this approach, is to segment and

² The relevant views are 2D because they are typically assumed not to contain information about the relative distances and orientations of the object’s visible surfaces in depth. For an argument that the problem of volumetric object recognition as premised on this assumption is ill posed, see Nakayama *et al.* 1995. This argument will play an important part in §§5-6 below.

³ This is not to deny that that top-down, object-specific knowledge may sometimes make contributions to lower-level visual processing that are not simply attentional in nature. See Kosslyn 1994, 235-36 and Kosslyn & Sussman 1995 for evidence that such knowledge may prime lower-level visual processing mechanisms for perception of an object’s spatial features via mental imagery. See Gregory 1997 for evidence that stored knowledge about faces and illumination conditions may dramatically affect perceived depth. For a contrarian defense of the modularity of early vision, see Pylyshyn 2003, chap. 4.

match 2D views of an object to the appropriate, stored aspect graph. (For discussion, see Bowyer & Dyer 1991, Van Effelterre 1994, and Tarr & Kriegman 2001.) Noë's claim that seeing a 3D object is a 'two-step' process – beginning perception of the object's 2D P-shape and ending with a set of expectations about how its 2D P-shape would vary with variation in one's point of view – seems clearly to recapitulate the process of object recognition as depicted by Koenderink and van Doorn.

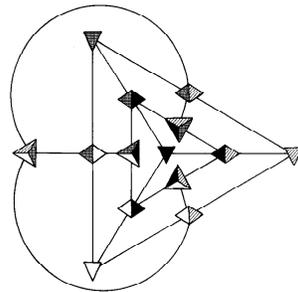


Fig. 1. The aspect graph or 'visual potential' of a tetrahedron. Nodes represent 2D aspects, while connections between nodes represent translations in 3D viewpoint space. (From Koenderink & van Doorn 1979 by kind permission of Springer-Verlag.)

One well-known difficulty with the aspect-graph-based approach to object recognition is that the number of distinct 2D views in the aspect graph for an ordinary, geometrically complex object may be quite large (Kriegman & Ponce 1990, Shimshoni & Ponce 1997). Koenderink (1984), for example, counts one view in the model of a globe, and 26 views in the model of a cube, but over 100,000 views in the model of a sculpture by the artist Giambologna. The computational challenge of matching or 'indexing' a given view to the model of an object comprising thousands of distinct views is formidable, especially if we don't assume that we already know which of indefinitely many possible different objects we are seeing. Terminological changes – e.g., for 'segmenting a 2D view of an object,' read 'seeing its P-shape,' for 'matching a 2D view of an object to a stored aspect graph,' read 'exercising implicit knowledge of the object's sensorimotor profile' – do nothing, of course, to lighten the computational load.⁴

Whether human object recognition actually utilizes internal representations of solid shape that have the properties of aspect graphs can only be decided by experimental psychology. For this reason Koenderink is quite careful to hedge claims about the psychological reality of the theory: 'It is up to visual psychophysics to determine the amount of structure actually sampled by human observers... we can at most hope to find a very coarse-grained version of the visual potential *in the mind*' (1984, 142). In the last twenty years, however, experimental evidence for the role of internal, aspect-graph-like structures in human object recognition has not been forthcoming. While aspect-graph-based modeling techniques have found application in computer vision and graphics, and while the methods Koenderink and van Doorn developed in order mathematically to represent the topological features that characterize unique object views have found application in research on human recognition (Tarr & Kriegman 2001), a wide variety of computationally less intensive

⁴ Noë would presumably be loath to characterize sensorimotor knowledge in such representationalist terms. I do not think, however, that it is possible to deflect the charge of what might be called *crypto-representationalism* by insistence on the 'implicit' or 'practical' character of the relevant knowledge.

and psychologically more plausible paradigms are currently being explored. (For a survey of recent viewpoint-dependent and viewpoint-independent approaches, see Tarr & Bülthoff 1999.) Noë's pronouncement in the passage quoted above that our visual abilities are sophisticated enough to manage the mathematically complicated sensorimotor profiles of 'animal bodies, plants, and so forth,' thus is not merely an *a priori* conjecture, it is quite probably false on empirical grounds.

For present purposes, the main points are two. First, the enactive account mistakenly assimilates the problem of visual shape perception to the problem of volumetric object recognition, thereby eliding the cognitive requirements of a relatively bottom-up (i.e., input-driven) process with those of a top-down process involving stored, object-specific knowledge. Second, the aspect-graph-based model of volumetric object recognition to which the enactive account assimilates visual shape perception is both computationally very expensive and, from the standpoint of current empirical research, psychologically unrealistic.

Although space does not allow adequate elaboration here, I would suggest that the mistake diagnosed above is actually symptomatic of a much more profound confusion about the constitutive role of *bodily action* in perception. According to Noë, 'to be a perceiver is to understand, implicitly, the effects of movement on sensory stimulation' (2004, 1). This is plausibly the enactive account's central constitutive claim. It epitomizes the account's competence model or 'task-level' analysis of spatial vision and, importantly, distinguishes the enactive account from numerous other action-oriented theories of visual perception. The problem is that the constitutive claim is false. In particular, there is no reason to suppose that in order to utilize spatial information in the light for purposes of skilful, spatially directed visuomotor action, biological visual systems need be able to predict or anticipate the way sensory stimulations, i.e., raw, unprocessed visual input, would vary with possible bodily movements.

The objection can be brought out initially by means of an example that O'Regan & Noë 2001 use to motivate their 'sensorimotor contingency theory' of perception (a forerunner of the enactive account):

Consider a missile guidance system allowing a missile to home in on an enemy airplane.... If the missile turns left, then the image of the target shifts to the right. If the missile slows down, the size of the image of the airplane decreases in a predictable way. The missile guidance system must adequately interpret and adapt to such changes in order to track the target airplane efficiently. In other words, *the missile guidance system is 'tuned to' the sensorimotor contingencies that govern airplane tracking. It 'knows all about' or 'has mastery over' the possible input/output relationships that occur during airplane tracking* (2001, 943, my emphasis).

The intended implication is that biological seeing is also subserved by implicit knowledge or understanding of the way sensory input would vary as a function of possible motor outputs. (Thus, 'according to the enactive approach, vision depends on one's knowledge of the sensory effects of, say, eye movements, for example, movements of the eye to the right causes a shift to the left in the retinal image' (2004, 25).) This is not, however, what the example would seem to show. Consider: Why is it plausible to suppose that the position of the airplane's image in the missile's sights represents or conveys information about the location of the airplane in missile-centric space? The answer, to a first approximation, is that the missile is able to make systematic use of the image in order to track and advance in flight toward the target. To function in this way, however, the missile needn't 'know all about' the various possible input/output relations. It needn't 'know,' e.g., that following the rule <turn toward the left when the image is on the left> will causally tend to bring it about that the image will move toward the right and, so, toward the middle of its sights. The missile

needn't represent – either implicitly or explicitly – the causal relationship between turning to the left on the output-side and 'foveating' the image on the input-side. It is enough for the missile successfully to exploit the image and, so, to fulfill its intended purpose that the relevant causal relationships between inputs and outputs do, as a matter of empirical fact, reliably obtain.

What confers a certain spatial representational content on the image, then, is its role in enabling image-consuming systems in the missile to perform their intended function of guiding the missile towards the location of its target. To perform this function, however, the relevant systems needn't have any expectations about how raw (unprocessed) image inputs would vary with different possible rocket engine outputs. They need only 'know' how to respond differentially to variations in the structure of the image. Similarly, what confers a certain spatial representational content on a pattern of retinal stimulation, or so I would argue, is its role in enabling stimulation-consuming systems in the living creature to initiate bodily movements targeted on or otherwise directed in relation to objects and surfaces in the distal, 3D environment. Variations in an object's shape and egocentric location, e.g., lead to variations in the structure of retinal stimulation, which, in turn, lead to variations in the way a perceiver scales her grip, rotates her hand, and extends her arm in order to grasp the object. (For neuropsychological evidence that the human visual system directly encodes incoming visual information using an action-oriented, egocentric spatial coding system, see Milner & Goodale 1995/2006, Rizzolatti *et al.* 1997, Andersen *et al.* 1997, Colby 1998, Colby & Goldberg 1999, Graziano & Botvinick 2002, and Jacob & Jeannerod 2003. For philosophical accounts premised on this idea, see Evans 1982, 1985; Peacocke 1992, chap.3; Brewer 1993; Campbell 1994; Clark 1997; Grush 1998, 2000; Gallagher 2005; and Briscoe *forthcoming*.) Biological seeing, of course, subserves a variety of other important purposes besides environmentally responsive visuomotor action – its contents are both 'directive' and 'descriptive' (Millikan 1996) – but this is plausibly its primary, i.e., evolutionarily central, function. In order to perform this function, however, a biological visual system needn't have any stored expectations or implicit knowledge about how raw sensory inputs would vary with different possible motor outputs.⁵ It is enough that the relevant 'laws of

⁵ This claim, I should note, is compatible with well-known evidence that the primate visual system exploits a 'corollary discharge' of the saccade command signal (SCS), i.e., a copy of motor instructions to the eye, in a predictive manner. Single-cell and fMRI studies of spatial remapping show that neurons at various stages in visual-processing exploit a corollary discharge of the SCS in order anticipatorily to shift their receptive field locations in the direction of an impending eye movement microseconds before its initiation (Colby *et al.* 1995, Merriam *et al.* 2003, Merriam & Colby 2005, Merriam *et al.* 2007). To simplify, a corollary discharge indicating an impending saccade 30° to the left in effect tells relevant neurons: 'If you are now firing in response to an item *x*, then stop firing at *x*. If there is now an item *y* in the region of oculocentric visual space that would be coincident with your receptive field after a saccade 30° to the left, then start firing at *y*.' Such putative remapping responses are strongest in parietal cortex and at higher levels in visual processing (V3A and hV4) and weakest at lower levels (V1 and V2). Further details needn't concern us here. For present purposes, the main point is that spatial remapping is a mechanism by means of which the visual system is able to create and maintain continuously up-to-date representations of distal, oculocentric object and feature locations. Hence, to the extent that remapping is properly regarded as predictive, the relevant predictions are to be understood at the level of *representational content*, i.e., changes in distal object and feature locations relative to the eye, not at the level of *representational vehicle*, i.e., correlative changes in proximal retinal stimulation. These remarks need to be developed with greater care than is possible here, but for relevant discussion of the proximal-distal distinction in the psychology of spatial vision, see Burge 1986, 24-39 and Millikan 2004, chap. 7.

sensorimotor contingency' do, as a matter of empirical fact obtain.⁶ But, if this is correct, then the enactive account seriously over-intellectualizes perception. It falsely elevates a causal, background condition of the possibility of spatially contentful visual experience to the level of knowledge required by the subject in order to enjoy such experience.

Noë, I should note, has responded to this charge. 'The charge of overintellectualizing,' he writes, 'is... answered, in two distinct ways. First, sensorimotor knowledge is knowledge of the way sensory stimulation varies as we move; it is knowledge that we share with non-linguistic creatures. Second, sensorimotor knowledge gets brought to bear in experience not in the form of judgment or belief or representation construction. The knowledge is practical, and we use it to gain and maintain contact with the world' (*forthcoming*, ms. 47). It seems clear, however, that neither answer actually addresses the gravamen of the case made here against the enactive account. The problem is not that the enactive account, so to speak, unwarrantedly 'conceptualizes' sensorimotor knowledge. The problem is rather that sensorimotor knowledge in the specific sense that Noë intends, i.e., knowledge of the way *proximal, physicalistically characterizable sensory stimulation* would vary with actual and possible bodily movements, is unnecessary to gain and maintain perceptual contact with the world.

Andy Clark has argued that any biologically plausible model in cognitive theorizing must comport with what he calls the '007 principle':

In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done (1989, 64).

In view of the foregoing discussion, it seems pretty clear that the enactive account violates the 007 principle. It stores in the modeled head what nature leaves to the world, namely, the 'laws of sensorimotor contingency.' Here are two more reasons to think that the enactive account violates Clark's constraint. The first has to do with what Mark Wexler 2004 calls *sensory aliasing*: 'Most of the time, current sensory states and motor commands can lead to many (usually infinitely many) future sensory states' (2004, 422). Moving sideways to the left while facing a solid object, e.g., not only leads to a corresponding shift of the retinal image to the right, but also to a deformation in the image that in general can only be predicted if one already knows the object's volumetric, 3D shape. More generally, the point is that it is normally impossible to predict or emulate changes in unprocessed visual stimulation on the basis of the motor output signal without assuming unrealistically detailed and complete knowledge of the spatial structure of the environment – not to mention the impending movements of other agents and objects, illumination conditions, etc. (See Wiener & Raab 2004 for related discussion.) If this right, however, then the worry is that the enactive account implicitly makes cognitive demands on the perceiver the satisfaction of which would, for all intents and purposes, make actually *seeing* the environment unnecessary.⁷

⁶ In this respect, they are akin to reliable environmental properties upon which evolved biological systems can 'piggyback' or use as 'external scaffolding' (Clark 1997, chap. 2).

⁷ Hence, it strikingly fails to follow through on the idea, notably endorsed by Noë himself, that perceptual systems can offload sensory processing by treating the world as 'its own best model' (Brooks 1991) or as an 'external memory' (O'Regan 1992).

Another reason to think that the enactive account violates the 007 Principle has to do with the role of conscious visual awareness in visuomotor action. According to Hurley & Noë *forthcoming*:

...it looks as if [an approaching] ball can be avoided by ducking because it looks as if ducking would make the situation look different in certain ways. Thus, when a ball looks as if it is moving towards one's face, it looks as if ducking would make the situation look different; as a result, it looks as if ducking would avoid the ball, and thus ducking to avoid the ball is enabled' (ms. 5).

This view does not have much *prima facie* plausibility. In particular, such anticipatory knowledge of the sensory effects of movements as is posited by Hurley and Noë seems neither necessary, nor sufficient for environmentally adaptive, real-time visuomotor action. It seems unnecessary, first, because there is abundant evidence that successful visuomotor action is possible without conscious visual awareness (Milner & Goodale 1995/2006; Clark 2001; Gazzaniga 1998; Koch 2004, chaps. 12 and 13) and, so, without knowledge of how an object or situation would *look* or *appear* different were one to move in certain ways. Well-known findings concerning subjects with profound visual form agnosia, e.g., show that rapid and fluent visually guided reaching, grasping, locomotion, etc., can be sustained despite pathological loss of normal abilities consciously to see the spatial properties of surrounding objects. Indeed, the possibility of such adaptive, visually guided in the absence of conscious visual awareness makes good sense from a phylogenetic perspective. As Gareth Evans writes, 'it seems abundantly clear that the evolution could throw up an organism in which... advantageous links [between sensory input and behavioral output] were established long before it had provided us with a conscious subject of experience' (1985, 387). But, if this is the case, then there is thus something fundamentally amiss with the suggestion that one's ability to avoid an incoming baseball is constitutively connected with knowledge concerning the way avoiding the baseball would make the situation *look* different. In general, knowledge of sensorimotor contingencies (understood here as regularities in the way 'looks' or 'visual appearances' undergo transformation with variation in one's point of view) would not seem to be necessary for visually guided action targeted on or otherwise directed in relation to an object.

It also seems clear that having such knowledge would not be *sufficient*. In particular, having such knowledge would not seem to explain or otherwise account for what Sean Kelly (2004) describes as our 'bodily readiness' to deal with an object's perceived features. It is hardly clear, e.g., how knowledge concerning the way a teapot's 'look,' i.e., its 2D perspectival shape and size, would vary with possible bodily movements might plausibly be supposed to enable one accurately to reach toward the teapot's handle, let alone to scale and adjust one's grip to its 3D shape and orientation. As Stephen Palmer (in criticism of the aspect graph approach) writes, 'Just from looking at an object, we generally feel that we know a great deal about its 3D structure, including how to shape our hands to grasp it and what it would feel like if we were to explore it manually. How can all this occur if all we have access to is structured set of 2D views?' (1999, 451). Noë often calls attention to similarities between seeing and touching an object – indeed he says that we ought to make touch our paradigm for perception – but, strikingly, the enactive account omits to provide any account of how seeing an object actually enables us to handle or otherwise interact with it. ('Visuomotor control' is not even in the index to *Action in Perception!*) In this respect and in notable contrast with other influential action-oriented theories

of visual perception – for an overview, see Mandik 2005 – the enactive account is decidedly *inactive*.⁸

3. Coin Tricks: Against the Phenomenological Claim

A traditional criticism of the sense-datum approach is that it makes an illicit inference from the *phenomenological* claim that a coin, when viewed from an angle, appears elliptical to the *perceptual* claim that this is the case because the subject *sees* something that actually is elliptical. Chisholm 1957 calls this the ‘sense-datum fallacy.’ (Also see Sellars 1968.) Notably, this criticism targets the sense-datum theorist’s interpretation of the phenomenological claim, but leaves the claim itself intact.

A more recent line of criticism due to Sean Kelly (2004) also concedes the phenomenological claim, but seeks to delimit its application. According to Kelly, it is possible to assume a certain behaviorally disengaged or detached perceptual attitude in which a tilted coin really does look elliptical. However, this is not the subject’s normal ‘engaged’ attitude. In the engaged attitude, the subject perceives the coin as round. (This is shown *inter alia* by how she acts in respect of the coin). For Kelly, the engaged attitude is primary. Hence, it is a mistake to explain what we see when we are in the engaged attitude in terms of what we see when we are in the disengaged attitude.

(In his ‘Real Presence,’ Noë offers in response to Kelly that ‘When I focus my attention on the coin’s shape, the elliptical appearance property that would be available to me as the focus of my attention were I to shift to the detached attitude, is present in experience only as a feature of the background. If this is right, then the opposition between engaged and detached attitudes is just a symptom of the figure/ground structure of conscious attention’ (*forthcoming*, ms. 11). Noë’s appeal to the figure/ground distinction from Gestalt theory seems far-fetched, however, since the putative shift of attention occurs *within* the region of the visual field bounded by the coin’s occluding edges, and, so, does not involve the coin’s relation to its background or the surrounding scene.⁹)

A substantially less accommodating line of criticism, carefully elaborated by Charles Siewert (1998) and A.D. Smith (2000a, 2000b), targets the sense-datum theorist’s phenomenological claim, i.e., the claim that an object’s visually apparent

⁸ I should emphasize at this juncture that I think that Noë’s heart is philosophically in the right place. I fully agree with him that perceiving is a skilful, bodily activity involving the animal as a whole and that we cannot hope adequately to understand perceptual representational content if we do not keep this in mind (see Briscoe *forthcoming*). At a lower level of generality, however, there is obviously much less about which we are in agreement.

⁹ In fact, Noë’s position on the relation between visual attention and visual awareness is quite ambiguous. ‘There is a sense,’ he writes, ‘... in which we move about in a sea of perspectival properties and we are aware of them (usually without thought or notice) whenever we are perceptually conscious. Indeed, to be perceptually conscious is to be aware of them’ (2004, 167). The problem is that this claim seems flatly to conflict with empirical findings concerning the phenomenon known as ‘change blindness.’ These findings strongly support the view that which details in a scene are consciously perceived – and so which can change without notice – is a function of visual attention. (For a review, see Simons & Rensink 2005a, 2005b). Noë, it should be emphasized, accepts this conclusion. ‘In general,’ he writes, ‘you only see that to which you attend. If something occurs outside the scope of attention, even if it’s perfectly visible (i.e., unobstructed, central, large), you won’t see it’ (2004, 52). Indeed, studies of change blindness play a prominent role in motivating his ‘anti-representationalism’ about visual perception (O’Regan & Noë 2001, Noë *forthcoming*). It seems clear, however, that Noë cannot have it both ways: he cannot maintain both that you are visually aware of P-properties even when you are not attending to them *and* that ‘you only see that to which you attend.’

shape is the shape of its perspective projection on the frontal plane. When we view the tilted coin, we see – in the most natural and straightforward sense of seeing – what looks like a disk that is partly nearer and partly farther away from us. We don't see something that looks (in either an epistemic or non-epistemic sense of 'looks') like a flat, upright ellipse.¹⁰ In general, the apparent shapes of the objects we see are shapes in depth. Thus, as Siewert nicely puts it, if there are any visual sense-data, then 'we will have to puff them out to three-dimensions' (1998, 227).

Ample empirical support for this objection is afforded by recent work in vision science. Indeed, the dominant 'information-processing' view in vision science is premised on the assumption that, to quote A.D. Smith, 'pre-conscious processes can extract 3D information from what is given to the eye, and can issue, as their *first* conscious upshot, in phenomenally 3D visual experience' (2000b, 492). Detailed discussion of the information-processing view is obviously impossible here, but key points to be made below can be motivated, I hope, by means of a few brief examples.

First, consider the role played by reflections in 3D shape perception. Recent psychophysical research has shown that, under real-world illumination conditions, reflections are a reliable source of information about surface curvature and texture (Takeuchi & Matsuoka 2002, Fleming *et al.* 2003, Fleming *et al.* 2004). In fact, reflections *alone* sometimes suffice to convey a vivid impression of 3D shape. Consider the following set of images from Fleming 2004:

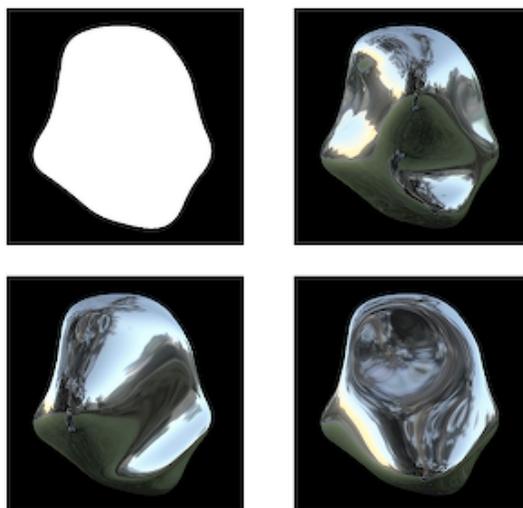


Fig. 2. Four images with identical silhouettes (P-shapes), but dramatically different apparent 3D shapes. Observers typically report having a vivid impression of 3D shape when viewing the three mirrored surfaces, even though the images contain no motion, stereo, texture, or shading. (From Fleming 2004 by kind permission of the author.)

In the figure, we see four objects with identical silhouettes (P-shapes), but dramatically different apparent 3D shapes. Note that the image of the silhouette by itself on the upper left seems flat or 2D. Also note that when viewing the mirrored surfaces in the other three images the only information about 3D shape is provided by *reflections* of the surrounding environment: the images contain no motion, texture, shading, etc. The key insight is that, under normal illumination conditions, i.e., the

¹⁰ See Austin 1962, 28 for a similar objection.

sorts of terrestrial illumination conditions in which the human visual system evolved, reflections on an object's surface are distorted in a way that systematically and measurably depends on the way the object's surfaces are oriented in depth ('shape-from-specularities'). These distortions appear as continuously varying, texture-like patterns, which Fleming and his colleagues call 'orientation fields'. They argue that the visual system can recover strong constraints on 3D shape from the orientation fields on a reflecting surface, much as it can recover strong constraints on 3D shape from the way texture patterns are compressed on a textured surface. An example of such 'shape-from-texture' is provided below (figure 3.):

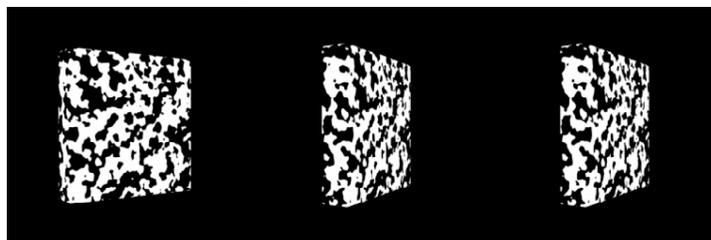


Fig. 3. A planar surface slanted at 30°, 60°, and 80° in depth. (From Fleming 2004 by kind permission of the author.)

In Figure 3, we see a uniformly textured, planar surface slanted at 30°, 60°, and 80° in depth. When the textured surface is rotated away from the frontal plane, the texture elements in the image are visibly compressed due to foreshortening. Such texture compression is often diagnostic of an object's 3D shape.

Other 'cues' to depth include occlusion (see §§5-6 below), motion parallax (relative motion), and shading. (For a more complete review and discussion, see Palmer 1999 and Bruce *et al.* 2003.) Perhaps the most important and widely studied source of depth information, however, comes from binocular disparity (stereopsis).¹¹ Binocular disparity arises due to small positional differences or 'disparities' between corresponding features in the two retinal images caused by the horizontal separation of the eyes. These disparities are used by the visual system to determine the relative depth between visible surfaces as well as their orientation (slant). A powerful example of the difference made by binocular disparity to our visual experience of the 3D world is provided by the neurobiologist Sue Barry. After living for decades without stereovision due to a congenital eye disorder, Barry underwent corrective therapy involving special eyeglasses. The results were dramatic. Here are two excerpts from her diary recently published by Oliver Sacks:

February 22: 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.

March 4: 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 3D space. The leaves didn't just overlap with each other as I used to see them. I could see the SPACE

¹¹ See Howard & Rogers 1995 and Schor 2004.

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 2006).

These diary entries vividly attest to the dramatic influence of binocular depth information on the spatial phenomenal character of our visual experience. Indeed, this is why *trompe l'oeil* paintings are traditionally supposed to be viewed with one eye shut. The visual illusion of three-dimensionality created by pictorial depth cues in the paintings is much enhanced when they are not in competition with the verdicts of stereovision.

For present purposes, there are two important points to take away from these familiar examples: First, there is ample empirical evidence for the view that an object's 3D shape is adequately specified by information in the light reflected from the object's visible surfaces and visually transduced by the eyes. Second, such shape-diagnostic information, having once been processed by the visual system, is not lost in our conscious visual experience of the object. Indeed, the examples above show that we can literally see difference made by their presence (and absence) in the light available to the eyes.¹² A coin's orientation in depth is as much part of the coin's visual appearance – how the coin looks when viewed *from here*, as Noë puts it – as is its color or texture (a point nicely articulated in Siewert 2006). It is important to emphasize that a more complete review of evidence for this view would require not only detailed explanation of the mechanisms by means of which the visual system utilizes individual sources of depth information, but also discussion of the way the visual system integrates or 'averages' simultaneously available and sometimes mutually inconsistent sources of depth information. (For an overview, see Cutting & Vishton 1995 and Bruce *et al.* 2003, chap. 7.)

We are now in a position to comment on Noë's recent claims about the perception of 3D shape in 'Real Presence.' Noë writes:

...what explains the fact that normal adult perceivers feel no inclination to take [a coin presented at an angle] to be elliptical on the basis of its visual appearance is the fact that they do have an abundance of countervailing information. Most of us know that coins are round, and we know that round, flat things change their perspectival shape as our spatial relation to them varies. As a matter of fact, there is considerable evidence that genuinely naïve perceivers – perceivers truly lacking the relevant countervailing information – would take round coins presented at an angle to be elliptical. Cheselden... describes the astonishment of a young boy who had undergone cataract surgery at the way a coin changed its shape as it moved; Helmholtz describes a similar case of a boy astonished at the changing shape of a locket (Noë, *forthcoming*,

¹² Perhaps Noë would reply that such 'mainstream' theorizing in cognitive science about how binocular disparity and other sources of depth-specific information are exploited by the visual system is fine so far as it goes, but it does not tell us anything, constitutively speaking, about *what it is* to see an object's spatial properties at the level of perceiver-environment interaction. 'No doubt perception depends on what takes place in the brain,' he writes, 'and very likely there are internal representations in the brain (e.g. content-bearing internal states). What perception is, however, is not a process in the brain but a kind of skilful activity on the part of the animal as a whole' (2004, 2). It is important to see, however, that such a reply would in no way serve to motivate the *specific* constitutive claim on which the enactive account is premised, i.e., the claim that in order to perceive an object's spatial properties it is necessary to understand how sensory stimulations caused by the object would vary with possible bodily movements. It is hard to see that such understanding plays any fundamental role in enabling skilful, visually based activity on the part of the animal as a whole if spatial information adequate for purposes of engaging in such activity is not only already available to the animal in the structure of the light reflected by the object, but is also demonstrably exploited by the animal's visual system.

ms. 5).¹³

Noë is surely correct to say that the reason that normal perceivers feel no inclination to take the coin to be elliptical on the basis of its visual appearance is that they have abundant *countervailing information*. The problem is that he seriously misrepresents both the character of that information and the normal perceiver's epistemic situation in visual perception. A normal perceiver does not see that the coin is round because she knows 'that coins are round, and... that round, flat things change their perspectival shape as our spatial relation to them varies.' Rather, we have seen that, under normal, informationally good viewing conditions, the coin's 3D shape and orientation are adequately specified by the structure of the light available as input to the visual system, i.e., such information as is provided by binocular disparity, reflections, shading, etc. The point is that the enactive account – in prescind from the 'bottom-up' contributions made to visual experience by such spatial information and by adverting instead to 'top-down' expectations about the sensory effects of movement – falsifies the phenomenological facts. A normal perceiver does not take a coin presented at angle to be round because she knows better than to take it to be elliptical. Elliptical simply is not in the running as a candidate for the coin's apparent shape.

(Hence, the perceiver's *expectations* with regard to what she would see were she to move in relation to the coin, to the extent that she has such expectations, are quite different than her expectations with regard to what she would see were she to move in relation to an upright ellipse. But the reason isn't, as Noë maintains, that the differences between the relevant sets of sensorimotor expectations are in some sense constitutive of the difference between her perceptual experiences in the two cases. It is just the other way around. Her sensorimotor expectations are different because what she *sees* is different.¹⁴)

Let us now turn to the claim that 'genuinely naïve perceivers... would take round coins presented at an angle to be elliptical.' In support of this claim, Noë adverts to the Cheselden case of 1728, a mainstay of philosophical reflection on what has come to be known as 'Molyneux's Question.' (In a letter sent to John Locke in 1688, the Dublin lawyer William Molyneux asked whether a blind man educated by touch to discriminate the shapes of 3D objects would visually recognize their shapes upon being gifted with sight.) There are good reasons, however, to be skeptical about phenomenological claims made on the basis of evidence about the post-operative experiences of cataract patients in the 18th century. As A.D. Smith has pointed out, excising cataracts at the time involved 'the complete removal or displacement of the lens of the eye' (2000b, 496). Hence, the post-operative experiences of patients would presumably reflect serious physical impairment of the visual organ. To make matters worse, the prolonged sensory deprivation suffered by the hitherto blind patients would, in all likelihood, have caused irremediable, neurally based perceptual deficits. David Hubel and Torsten Wiesel, e.g., in well-known animal studies showed that, when binocular vision is averted by a congenital defect or by surgical intervention,

¹³ Noë misremembers the details of the two cases he mentions here. It is Helmholtz who in his *Physiological Optics* (1909/1925) remarks on a patient's surprise at the way a coin appears to change shape when rotated and Cheselden (1728) who describes a young boy's astonishment that his father's likeness should fit within the confines of his mother's locket.

¹⁴ Compare Merleau-Ponty: 'What, then,... is seeing a cube? It is, say empiricists, associating with the actual aspect of the drawing presented, a set of other appearances, those that it would present at closer quarters, from the side, from various angles. But, when I see a cube, I do not find any of these images in myself; they are the small change of a perception of depth which makes them possible, but which does not result from them' (1945/1962, 264).

binocular cells in the visual cortex fail to develop (Hubel & Wiesel 1963, Wiesel 1982). In consequence of such developmental brain damage, the functioning of later vision, as Shaun Gallagher comments, 'will be limited and quite different, not only from the normal adult, but also from the newborn infant. In neurophysiological terms, the neonate is not equivalent to the Molyneux patient' (2005, 166).

The upshot is that there is little reason to suppose that the reports about the post-operative experiences of cataract patients tell us anything about 'genuinely naïve' perception. Indeed, contemporary medical studies suggest that what patients first 'see' after cataract surgery is not an array of 2D forms, but rather a meaningless blur of lights and colors (Gregory & Wallace 1963, Sacks 1995). (Noë, I should note, is forgetful that he himself once observed that cataract surgery on the congenitally blind 'does not in fact restore sight. In none of the well-known cases, from the famous Cheseldon [*sic*] case down to the present day, does surgery result in vision' (Noë 2002, 68).) Moreover, if it can be assumed that neonates and very young infants *are* genuinely naïve perceivers, then, as Smith 2000b and Gallagher 2005 convincingly argue, there is fairly persuasive developmental evidence that the intrinsic or 'original' spatial phenomenal character of visual experience is 3D. From first perception, human beings experience objects as arrayed in 3D space.

Some of the main findings about the visual abilities of neonates and very young infants, based on standard preferential looking measures, include the following:

1. Neonates visually perceive *shape* as invariant across changes in retinal stimulation caused by changes in object orientation, i.e., shape constancy is present at birth. There is evidence both that neonates detect and respond to changes in an object's slant in depth, and also that they respond to an object's real shape, regardless of its slant (Slater & Morison 1985).
2. Neonates visually perceive *size* as invariant across changes in retinal stimulation caused by changes in object location, i.e., size constancy is present at birth. Thus, in one experiment, neonates differentiated between the size of a large cube and a less distant small cube despite the similarity of their retinal sizes (Slater *et al.* 1990).
3. Neonates selectively fixate a 3D stimulus in preference to a photograph of the same stimulus, even under monocular viewing conditions and even when the major depth cue is motion parallax (Slater *et al.* 1984).
4. Infants as young as eight weeks perceive 3D object shape when shown kinetic random-dot displays (Arterberry & Yonas, 2000).

Weighing similar evidence about neonatal visual abilities, Smith draws the following conclusion: 'the idea that sight must acquire a 3D significance has a probability of approximately zero.... Indeed, in the absence of both argument and developmental perspective it is difficult even to *make sense* of the claim that three-dimensionality is not 'original' to sight' (2000b, 502-3).¹⁵ If this is right, however, then there is simply no empirical substance to Noë's claim that 'genuinely naïve' perceivers would visually experience a round coin presented at an angle as elliptical.

¹⁵ This is not to deny, of course, that human neonatal visual abilities are quite limited in comparison with those of more mature perceivers. There is evidence, e.g., that stereopsis, for example, becomes fully operative only around the end of the fourth month of life (Atkinson & Braddick 1989).

4. Noticing P-Properties as Make-Perceive

We have seen that the real objection to the phenomenological claim is not, as Sean Kelly (2004) suggests, that in order to see the elliptical 'look' of a coin presented at an angle we must adopt a certain detached perceptual attitude. This objection, it is now clear, concedes far too much. The coin does have an elliptical P-shape – this we know from the laws of linear perspective – but the coin looks round (and slanted in depth). The real objection to the phenomenological claim, we have seen, is that the coin simply does not have an elliptical 'look.'

What about the phenomenologically weaker claim that, in seeing the coin, we are *able to see* its elliptical P-shape? After all, even if an object's visually apparent shape is fully 3D, might it not be the case that we also see its 2D P-shape? The enactive account, of course, is premised on this supposition. 'P-properties,' Noë writes, 'are themselves *objects of sight*, that is, things that we see. They are visible' (2004, 83). And also: 'The crux is that P-properties are not merely visible qualities, such as shape and size. They are *looks* of things, their visual appearances' (2004, 84).

The burden of this sections will be to show that this supposition is mistaken. When we see an object under standard perceptual circumstances, we see its visually apparent 3D shape, i.e., the orientation of its visible surfaces in depth, but we do not also see its 2D P-shape. In general, P-shapes are not properties that 3D objects under standard perceptual circumstances can literally be seen to instance. (This, however, as I point out below, is not to deny that P-shapes are sometimes visible when we see certain pictorial representations of objects.)

Perhaps the most straightforward objection to the claim that in seeing an object under standard perceptual circumstances we are able to see the shape of the patch that would occlude the object on a plane perpendicular to the line of sight is that both the plane and the patch are hypothetical, i.e., counterfactual. Presuming that we do not in any literal sense perceive hypothetical objects (we do not *inter alia* stand in any causal relation to them), and that we do not see properties without seeing the objects that appear to instance them (there are no 'disembodied' properties), it follows that we do not see an object's P-shape. We no more see the shape of the patch that *would* occlude a coin on a plane perpendicular to the line of sight when viewing the coin, than when viewing the sky at night we see the lines that *would* connect the stars in Orion's belt.

J.J. Gibson provides another objection to the claim that we are able to see an object's P-shape:

What one becomes aware of by holding still, closing one eye, and observing a frozen scene are not visual sensations but only *the surfaces of the world that are viewed now from here*. They are not flat or depthless but simply unhidden (1979, 286).

Drawing in perspective does depend on viewing in perspective, it is true, but this only means that drawing requires the learner to notice the edges of the layout confronting him, especially the occluding edges.... What we loosely call an outline in a picture refers to the outer edges of the face of an object.... I am saying that edge perspective is a fact, whereas patchwork perspective is a myth. One can learn to view the former but not to see the latter (1979, 287).

Gibson's point in these passages, I take it, is that an object's occluding edges are perceptibly oriented *in depth*: 'the surfaces inside the field of view... are not a depthless patchwork of colors, for they have the quality that I [call] slant' (1979, 196). If this is right, then to attend to an object's occluding edges, to limn them with the line of sight, so to speak, is not in itself to see its P-shape. The occluding edges of the objects we see, in contrast with the contours of their P-shapes, are oriented in 3D

space. This objection to what Gibson calls ‘patchwork perspective,’ incidentally, bears on Noë’s proposal that ‘the plate looks elliptical to me because, to indicate its shape, I can (and indeed, in some sense, *must*) move my hand in a characteristic manner. That is, to experience a thing as elliptical is precisely to experience it as occupying a particular kind of region in one’s egocentric, sensorimotor space’ (2004, 89). The problem with this proposal, we are now in a position to see, is its assumption that what I succeed in indicating when I move my hand in an ellipse is the plate’s 2D P-shape on the frontal plane and not simply the plate’s occluding edges in 3D space.

I have been providing reasons to resist the claim that when we are presented with a 3D object under standard perceptual circumstances we are able to see its P-shape. This, however, is not to deny that we are able to see an object’s P-shape in a perspectively accurate picture or photograph. Indeed, Richard Wollheim has argued that it is essential to our visual experience of a pictorial representation that it comprise both a ‘recognitional’ and a ‘configurational’ aspect (1980, 1998). In seeing a pictorial representation, we visually recognize what is pictured in seeing the physical configuration of the pictorial surface. Wollheim calls this phenomenon ‘seeing in’ (in deliberate contrast with ‘seeing as’). Thus we may see the 3D shape of a cat *in* the 2D patch of photo-pigment corresponding to the cat on the surface of a photograph. It is a necessary feature of ‘representational seeing,’ of the seeing appropriate to the perception of a picture *as* a picture, that we be able to attend both to the object represented by the picture and to the physical medium/vehicle of representation. Wollheim refers to this requirement as the ‘two-fold thesis.’

In *Action in Perception*, Noë suggests that Wollheim’s analysis of representational seeing can be extended in order to explain straightforward perception of 3D spatial properties: ‘Just as it is the case that you see a picture, and, in seeing the picture, you see what the picture depicts..., so I want to suggest that we see the elliptical perspectival properties [of a plate] and, in seeing them, we see the plate’s circularity (and so in that sense see the circularity in the elliptical perspectival properties)’ (2004, 241, n. 8). The trouble, if the foregoing discussion is correct, is that there is nothing plausibly analogous in the world to a 2D pictorial surface – or to a 2D patch of photo-pigment or paint – in straightforward perception of a round plate. In seeing the plate, the configurational aspect of what Wollheim calls representational seeing is missing from our experience. Hence, the notion of ‘seeing in’ does not find legitimate application.

Noë works hard to motivate the claim that P-shapes are properties that 3D objects, and not just pictures of 3D objects, can literally be seen to instance. I think that our phenomenological intuitions, however, are easily distorted by reflection on the sorts of examples favored by Noë and by sense-datum theorists in general. The reason is that it is easy to notice – in a sense that I shall presently explain – the simple and regular P-shape of a coin, or a tabletop, or a globe, while it is exceedingly difficult to notice the intricate and irregular P-shape of an oak tree, or a rose bush, or an abstract sculpture by Frank Stella. I want to claim that noticing the P-shape of a coin, to take the simplest case, is so easy, in fact, that we mistake the act of noticing its P-shape for the act of *seeing* its P-shape. Having made this mistake, we then suppose that there is such a thing in general as ‘seeing an object’s P-shape.’ And once our phenomenological intuitions are distorted in this way, it is then possible for us to take seriously the suggestion that being able to see an object’s 3D shape in some sense depends on being able to see its 2D P-shape.

What, then, do I mean by ‘noticing’ an object’s P-shape? I can begin by explaining what I *don’t* mean. I don’t mean making the visually based judgment (or forming the visually based belief) that the object would project a certain 2D shape on a plane perpendicular to the line of sight. Noticing that a coin viewed from an angle has an elliptical P-shape is not like noticing that Alice’s dress is new because the price

tag is still on it. My – admittedly speculative – proposal is rather the following: Noticing an object's P-shape involves *an exercise of visual imagination guided by seeing*, an act of what one might call 'make-perceive.' In general, one engages in make-perceive when one projects or 'superimposes' a mental image on a certain region of the visually perceived world. (Gosselin & Schyns 2002 call such hybrid experiences involving both top-down, internally generated and bottom-up, externally generated components 'superstitious perception.')

To make-perceive an object's P-shape is imaginatively to construct and superimpose the object's P-shape on the region occupied by the object in one's field of view.

This proposal, I should emphasize, isn't at all outlandish. Indeed, certain examples of make-perceive are quite familiar. Thus, as Rick Grush observes, one might use a mental image in order to decide where in egocentric space a vase should be placed in order best to obscure a picture on the desk (2004, 390). Similarly, when deciding how to arrange the furniture in a new home, one may begin by visualizing how things would look were one to place an armchair in a certain corner, or were one to hang a painting on a certain wall.¹⁶ It seems clear that one does not in any literal sense *see* the armchair in the empty corner or the painting on the unadorned wall.

The claim I wish to make, then, is that to notice an object's P-shape is knowingly or unknowingly to imagine the shape of the 2D patch that would occlude the object while simultaneously viewing the object. One does this by superimposing an off-line mental image of the object's P-shape (perhaps created in short-term visual memory) on top of its on-line 3D shape. The effect, so to speak, is to trace the object's P-shape on an imaginary pane of glass.¹⁷

It is important to emphasize that the ease with which one make-perceives an object's P-shape depends, among other things, on the object's 3D shape and orientation. Notably, one's ability to make-perceive the simple and regular P-shape of a coin, or a tabletop, or a globe is typically effortless. Hence, the tendency, or so I have argued, to conflate the act of make-perceiving the P-shapes of these objects with the act seeing of their P-shapes. And, hence, as J.L. Austin points out, the sense-datum theorist's 'constant, obsessive repetition of the same small range of jejune "examples"' (1962, 3).

According to this model, it is not necessary actually to see the P-shape that would occlude an object in order to notice it. One can make-perceive it instead. When seeing the object's 3D shape, attention is directed to the object itself. When make-perceiving its 2D P-shape, by contrast, attention is directed to properties of an internally constructed mental image superimposed on the region occupied by the object in one's field of view. For this reason, the less 'distracting' depth information about the object's surface orientation(s) is afforded by texture, shadows, reflections, etc., the easier it is to make-perceive its P-shape. Thus, when children learn to draw an object in perspective, they are sometimes taught to view the object with one eye shut or slightly to defocus their gaze, the intended effect in both cases being to suppress depth cues that might otherwise interfere with the task. Of course, in the absence of reliable information about depth, a 3D object *will* sometimes look flat or 2D (and vice versa, see Ames 1951, King *et al.* 1976). It seems clear however that the possibility of such visual illusion under informationally impoverished viewing

¹⁶ A related case discussed by Wollheim (1980, 218) is Leonardo da Vinci's advice to the aspirant painter to look at the patterns on water-stained walls and imaginatively to attempt to 'see' battle scenes and mysterious landscapes in them. Wollheim's observation that such imaginative-visual experience is marked by a certain 'indifference' between attending to that which is imaginatively projected (the battle scenes or landscapes) and attending to the world (the actual patterns on the wall) is directly relevant to the proposal made in this section.

¹⁷ This metaphor was suggested in conversation with Ruth Millikan.

conditions by itself provides no support for the sense-datum theorist's claims about what we see under normal, informationally rich viewing conditions. No more, say, than the possibility of mistaking a cow for a horse on a dark night provides support for the claim that cows look like horses in broad daylight. Two *different* visual experiences are at issue.

If the foregoing proposal is on the right track, then we do not, as Noë suggests, see an object's 3D form in its 2D P-shape. Just the opposite is the case: we sometimes make-perceive its P-shape in its 3D form. I encourage the reader to choose a well-illuminated and irregularly shaped 3D object – i.e., *not* a coin, or a tabletop, or a globe – and to monitor carefully what you actually do in purposing to notice the object's P-shape. Try to notice the object's P-shape *without* 'tracing' its boundaries in imagination. I do not think you will succeed.

I should emphasize that, if the model finds empirical validation (and one Popperian virtue of the model is that experimental studies could in fact be devised to test it), then it dispenses with a question that has preoccupied philosophers at least since Locke: namely, the question of how an object can visually appear to instance *both* a certain 3D shape and a certain 2D perspectival shape at the same time (see Schwitzgebel *forthcoming*). Indeed, it is this putative tension that leads Noë to claim that perceptual content is 'radically ambiguous' (2004, 34) and that visual experience is rife with 'dueling phenomenological facts' (2004, 78). The enactive account is intended to relieve this tension, to reconcile the fact that 'perceptual experience represents the world as voluminous... *and* that you cannot see anything in front of which an opaque surface is placed' (2004, 78). What I hope to have shown in the foregoing is that there is no such phenomenological conflict in visual experience. We perceive an object's 3D shape, i.e., the orientation of its visible surfaces in depth, but we *make-perceive* its 2D P-shape. We do not visually experience the object as both voluminous and flat at the same time.

5. Depth and the Perceptual Organization of Visible Surfaces

When we view the perspective projection of a scene in a painting or photograph, we see the projected scene in what Wollheim calls the configuration of the pictorial surface. The term 'configuration,' however, is ambiguous. On the one hand, we see the projected scene in a certain unstructured, 2D array of colored points or *pixels* on the projection plane of the pictorial surface. In this sense, the configuration of the pictorial surface is something that could be adequately represented by means of assigning a certain hue, saturation, and brightness to every pixel in the picture. Call what would be represented by such a total assignment the '**photometric-configuration**' of the projection plane. On the other hand, we see the projected scene in a certain structured, 2D array of P-shapes (Ruskin's 'flat stains of colour'). In this sense, the configuration of the pictorial surface is something that could be adequately represented by means of outlining the visible contours of objects pictured in the image. Call what would be represented by such an outlining of P-shapes the '**perspective-configuration**' of the projection plane. (In the object recognition literature, this would be a complete segmentation of all discrete object boundaries in the corresponding 2D image.)

We can also speak, of course, of the configuration of the distal scene beyond, i.e., on the far side, of the projection plane. In viewing Seurat's *Un dimanche après-midi à l'Île de la Grande Jatte* (figure 4), e.g., we perceive an organized, coherent array of objects and surfaces in 3D space. In particular, we see that certain objects in the scene are closer than others, i.e., closer to the pictorial point of view, and, in seeing this, we see how they occlude objects and surfaces that are farther away. The scene, in other words, has a certain 3D perceptual organization substantially characterized by

relations of occlusion between objects and surfaces arrayed at different distances in depth. Call this distal, 3D organization the ‘**depth-configuration**’ of the projected scene.

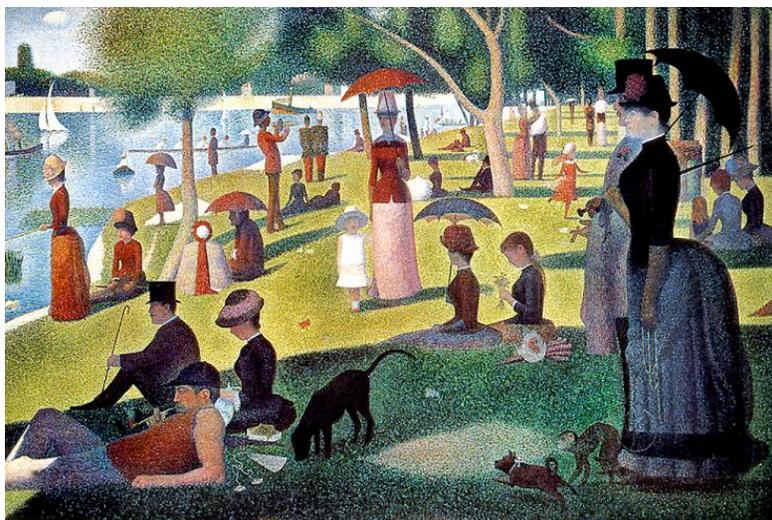


Fig. 4. G. Seurat, *Un dimanche après-midi à l'Île de la Grande Jatte* (1884-1886), Art Institute of Chicago.

According to the enactive account, the perspective-configuration of the projection plane is ‘perceptually basic.’ When we see an object, what we see, in the first instance, is its 2D P-shape. It is only in consequence of successfully matching its P-shape to a certain stored sensorimotor profile that we are able to perceive its 3D shape in depth. The enactive account, however, takes perceptual awareness of the perspective-configuration of the projection plane for granted. I.e., it provides no explanation of how the subject on the basis of information available to the eye perceptually individuates or segments 2D regions on the projection plane that correspond to the contours (occluding edges) of 3D objects in the distal scene. E.g., when Noë discusses the example of the tilted coin, he simply assumes that we are able to segment the elliptical 2D P-shape of the coin as figure from background.

The trouble is that there is nothing intrinsic to the photometric-configuration of the pictorial projection plane – or, crucially, the retinal image – that determines a unique perspective-configuration. As David Marr writes:

Regions that have ‘semantic’ importance do not always have any particular visual distinction. Most images are too complex, and even the very simplest, smallest images... often do not contain enough information in the pure intensity arrays to segment them into different objects (1982, 270).

Edges that ought to be significant are either absent from an image or almost so... and the strongest changes are often changes in illumination that have nothing to do with meaningful relations in a scene (1982, 272).

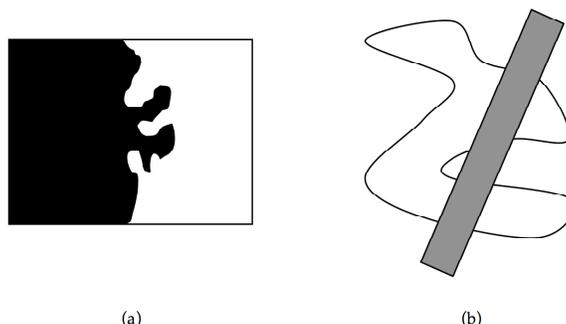
In the present context, the point is that were we to segment the image the projected by a typical, real-world scene, e.g., a densely cluttered forest setting, solely on the basis of the most abrupt luminance intensity transitions in the image, the segmentation achieved would not reliably correspond to the contours of the objects visible in the scene. It would not match the perspective-configuration of the

projection plane. Many objects would be ‘broken up’ by the patterns of light and shadow in the image.

An equally important consideration is the fact of occlusion. It is a consequence of inhabiting a 3D world that near objects very often partially hide objects farther away. (Again, think of a typical forest setting with crisscrossing branches and overlapping leaves.) A medium-sized region on the projection plane, in normal environmental contexts, will typically contain points projected by multiple surfaces at different distances in depth from the viewer. Which points in the 2D photometric-configuration map onto which object surfaces in the 3D depth-configuration is not something that in general can be read directly off the intrinsic properties of the former. Since this mapping, however, is what determines the perspective-configuration of the projection plane, it again follows that the perspective-configuration of the projection plane is extrinsic to its photometric-configuration.

In what follows, I argue that which perspective-configuration, which structured array of P-shapes, is perceived on a projection plane psychologically depends on perception of the depth-configuration of the projected scene. Our ability perceptually to individuate P-shapes on a 2D projection plane depends on our more basic ability perceptually to individuate objects and surfaces in the 3D projected world.

This point can initially be brought out by looking at some uncomplicated examples of the role played by depth perception in perceptual organization. Consider figures 5a and 5b below:



Figs. 5a and 5b. Depth and perceptual organization.

In figure 5a, our perception is bi-stable: which region of the image we see as figure and which as ground reverses. The bi-stability in the perception can be explained in terms of the bi-stability in the perceived order of depth (Nakayama *et al.* 1995). When perceived as closer, the white region is seen as an occluding surface, with the black region continuing behind. With a perceptual re-ordering of depth, ‘ownership’ of the border between two regions flips, and the reverse occlusion relation is perceived to obtain. In figure 5b, we effortlessly and automatically see what appears to be a white surface partially occluded by a grey strip. I.e., a grey strip appears to be closer in depth and to hide parts of a white surface that complete behind it. Our visual impression is not of four unrelated image regions on the same plane of depth.

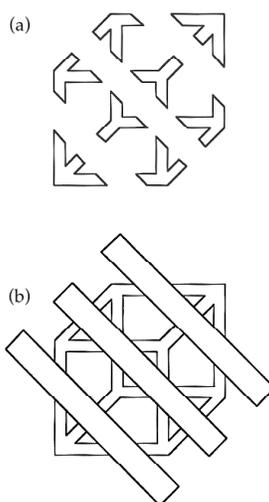
These simple examples serve to illustrate the point that the perceptual organization of visual experience crucially depends on preconscious processing of *depth relations* between surfaces and other perceptual units. The structure of visual experience reflects the physically significant spatial structure of the distal, environmental scene – what I have called its ‘depth configuration’ – rather than the photometrically significant properties of the scene’s 2D plane projection. In the next

section, using more sophisticated examples, I shall provide more evidence that our ability to impose a certain perspective-configuration on the photometric-configuration of the projection plane is psychologically reliant on our ability to perceive the depth-configuration of the 3D world beyond the projection plane.

6. Surface Representation and Object Recognition

A forceful and influential argument for the perceptual priority of surface layout in depth has been made by Ken Nakayama and his colleagues (1995). Such surface representation, they argue, 'is a necessary intermediate form of perceptual representation, one that forms an appropriate foundation for other visual functions – object recognition, object manipulation, and locomotion' (Nakayama *et al.* 1995, 25).¹⁸ In what follows, I focus on their case in relation to object recognition.

Consider first figure 6a below. When we look at the figure, our visual impression is of eight, separate, 2D image fragments on the *same* plane of depth. We may with prolonged inspection imagine the fragments as joining to form a Necker cube, but that is not what we in the first instance see.



Figs. 6a and 6b. Occlusion, perceptual grouping, and recognition. (From Kanizsa 1979 by kind permission of the Greenwood Publishing Group.)

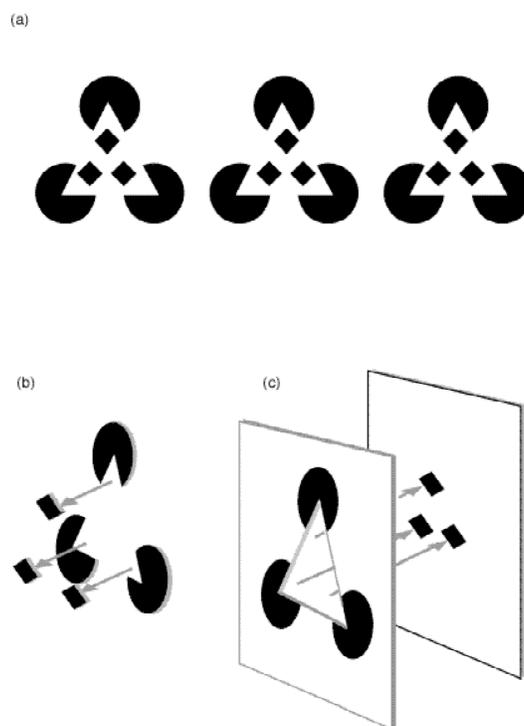
Next consider figure 6b. The visual impression of a 3D Necker cube is now quite vivid. There is no tendency to see eleven separate, co-planar image fragments, i.e., the eight original image fragments plus three additional strips. It is important to recognize that the perception of *occlusion* in the image is responsible for this effect (Kanizsa 1979). Since the surfaces of the three, co-planar strips in 6b are perceived as in front, they 'own' the borders between them and the surface regions that are perceived as behind. (The relevant pictorial depth cues here are the 'T-junctions' in the image, which in accordance with the so-called 'law of generic views',¹⁹ specify the

¹⁸ For recent findings in support of this thesis, see Johnson & Olshausen 2005.

¹⁹ The law of generic views asserts that the visual system when confronted with more than one possible surface interpretation of a scene assumes the viewer is seeing the scene from a 'generic,' i.e. a non-accidental, vantage point. The most probable interpretation of 6b is that of

'caps' on the junctions as in front and the 'stems' as behind.) And since surfaces that do not own borders are effectively unbounded, they can connect to other unbounded surfaces and amodally complete behind their occluders (Nakayama *et al.* 1995). The point of relevance is that there is nothing intrinsic to the photometric-configuration of 6b that dictates how 2D image fragments in the figure are either to be grouped or segregated. Visual awareness of the 3D cube depends rather on the perception of occlusion and, so, on the perception of the depth configuration of visible surfaces pictured in the image.

Not surprisingly, similar results obtain in connection with stereoscopic depth. By switching left and right images in a stereogram it is possible to reverse perceived depth in a scene without changing the total amount of monocular visual information available to the two eyes. Such reversal of depth can cause fundamental changes in the perceived 3D layout and grouping of surfaces. This is illustrated by figure 7 below:



Figs. 7a-7c. Large-scale change in surface representation with reversal of perceived depth. (From Fleming & Anderson 2004 by kind permission of MIT Press.)

This figure is adapted from Takeichi *et al.* 1992. Each image in 7a consists of a Kanizsa 'illusory' triangle and three diamonds. Cross-fusing the stereopair on the left side of 7a has the effect that the diamonds appear nearer in depth to the observer than the triangle and its inducers. The result is shown in 7b. Notably, the triangle in 7b appears as a *figure* in front of three (amodally completed) disks. The perceived

a partially occluded cube rather than some accidental arrangement of eleven discrete, 2D surfaces. See Hoffman 1998 for discussion.

disparities in depth can be inverted – again, without changing the total amount of monocular visual information available to the two eyes – by switching the left and right eye’s views. This is achieved by cross-fusing the stereopair on the right side of 7a. The effect now is that the diamonds appear to be behind the plane of the inducers. The result is shown in 7c. As Fleming & Anderson 2004 point out, ‘this simple inversion leads to a change in surface representation that is more complex than a simple reversal in the depth ordering of the perceptual units’ (1286-1287). When the diamonds recede, they ‘drag’ their background back with them. Notably, the triangle now appears as a *hole* in a surface through which the observer can see the diamonds.

There are two main points. First, the perspective-configuration of 7b, i.e., its segmentation into distinct P-shapes, is clear significantly different than the perspective-configuration of 7c. The example thus illustrates the way in which the perceptual organization of visual experience varies dramatically with changes in apparent (stereoscopic) depth. Second, prior to the process of object recognition – or prior to the process matching a 2D P-shape to a stored sensorimotor profile – an object must be segmented, i.e., its boundaries must be isolated, from the rest of the image. What these and other examples show is that such segmentation can crucially depend on perceived depth relations between visible surfaces in the image. It cannot in general be achieved solely on the basis of intrinsic, 2D image features. Thus Nakayama *et al.* write:

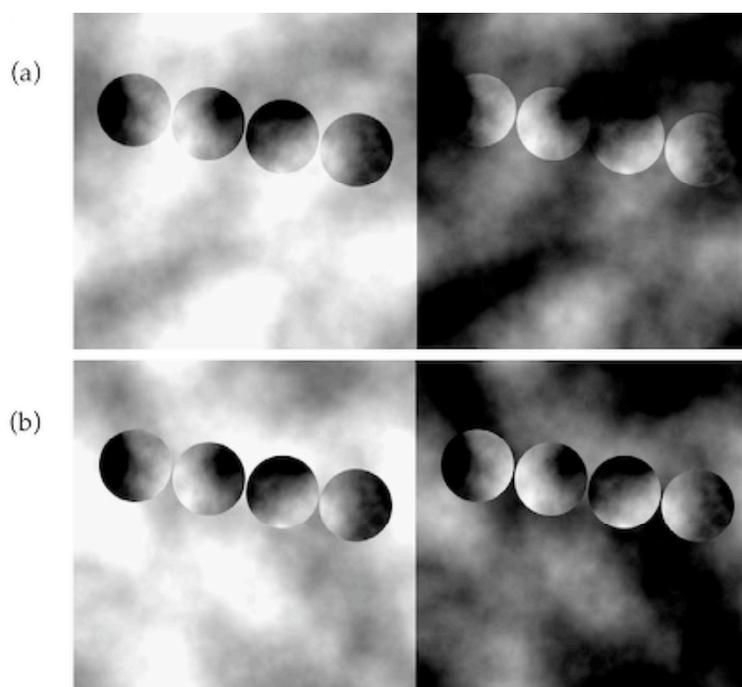
[O]ur perception of recognizable objects can be dramatically influenced by visual surface representation.... Rather than being used to represent the internal three-dimensional structure of the objects themselves, depth has a more important role: *it determines what pieces of an image actually comprise the object to be recognized.* Depth is needed to parse objects into wholes or parts, to determine what in an image actually constitutes the parts and boundaries of a single object. In other words, depth dictates perceptual grouping and perceptual segmentation (1995, 13-14, my emphasis).

The relevant implication in the present context is that perception of an object’s 3D shape cannot in general proceed from prior visual awareness of the object’s P-shape, because visual discrimination of the object as figure from ground and, so visual discrimination of its P-shape, depends upon prior perception of the way visible surfaces in the scene are arrayed in depth.²⁰

The enactive account, like the classical sense-datum theory of perception, is premised on the assumption that ‘perceptual awareness is, in the first instance, an awareness of perspectival qualities’ (Noë 2004, 165). The phenomenological and experimental evidence reviewed in this section amply shows that this assumption is false. While some of the points made above were made in connection with perception of depth in pictures, they apply with full vigor to perception of depth in actual scenes. The idea integral to the enactive account that we are visually aware of the shape of the 2D patch on a plane perpendicular to the line of sight that would occlude an object presupposes that the relevant patch has already been segmented as figure from ground. What we have seen is that such segmentation strongly depends upon prior perception of the 3D structure or ‘depth configuration’ of the visible scene. If this is right, then it provides vivid demonstration that P-properties cannot play the perceptually basic role conferred on them by the enactive account.

²⁰ This last claim, I should point out, would be contested by many researchers in the visual object recognition field (see Edelman 1999 for relevant discussion). There are many computational approaches that pursue segmentation solely on the basis of 2D image features. To the best of my knowledge, however, no such approach works robustly, i.e., reliably well across a wide variety of scenes in ecologically realistic viewing conditions.

Although space does not permit adequate discussion, I should mention in concluding this section that Noë's discussions of *apparent lightness* are also enfeebled by an inattention to the role of depth. Noë rightly observes that 'The way a thing looks with respect to... brightness depends not only on viewing geometry and lighting, but also on the chromatic properties of surrounding and contrasting objects' (2004, 125). He omits to mention, however, that the *albedo* or apparent lightness of a surface is dependent not only on its perceived adjacency to other surfaces in the retinal image – and, by extension, its perceived adjacency to other surfaces in what I have been calling the perspective-configuration of the projection plane – but also on its perceived position relative to other surfaces and orientation in 3D space (Gilchrist 1977, 1980; Rock 1983, chap. 9; Knill & Kersten 1991; Anderson 1999; Anderson 2003). Indeed, it has been demonstrated that the apparent lightness of a region of the visual field can be switched from white to black by varying its apparent depth relative to other regions. A quite striking example of this effect is provided by figure 8a below from Anderson & Winawer 2005:



Figs. 8a and 8b. (a) The corresponding disks on the light and dark surrounds are physically identical, but the disks on the light surround appear as dark objects seen through semi-transparent dark haze, whereas the disks on the dark surround appear as light objects seen through semi-transparent light haze. (b) Rotating the surround 90° destroys both the perception of transparency and the lightness illusion. (From Winawer & Anderson 2005 by kind permission of the Nature Publishing Group.)

In figure 8a, the corresponding disk regions on the light and dark surrounds are physically isoluminant (pixel for pixel), but the disks on the light surround are perceived as uniformly dark objects behind intervening light haze, whereas the disks on the dark surround appear as uniformly light objects behind intervening dark haze. Crucially, this lightness illusion is to be explained by the way the difference in the lightness of the surround modulates the perception of transparency inside disk regions. More precisely, whether a disk in the example appears as light or dark depends on the way the visual system uses the lightness of the surround in order to

segment the contrasting luminances that define the texture inside the occluding boundary of the disk into discrete light and dark surfaces *at different distances in depth*. The process whereby the visual system achieves perception of transparency by segmenting multiple surfaces in depth along a single line of sight is known as ‘scission’ (Koffka 1935). For present purposes, the details of this process needn’t concern us (but for careful explanation, see Anderson 2003 and Fleming & Anderson 2004). What matters is that when the surround is light, the grey regions, i.e., intermediate luminances, in the texture are interpreted as dark matter seen behind semi-transparent light matter. When the surround is dark, the inverse relationship in depth between the light and dark surfaces is perceived to obtain, and the grey regions in the texture are interpreted as light matter seen behind dark matter. That segmentation processes involved in the perception of transparency play a critical role in this lightness illusion is shown by the fact that rotating the surround in the figure 90° not only contravenes the photometric and geometrical conditions needed to induce perception of transparency, but also destroys the lightness illusion. This is illustrated by figure 8b, in which the disks appear neither uniformly white, nor uniformly black.

What this last example demonstrates is that the way a thing looks with respect to brightness depends not only on viewing geometry and lighting conditions, and not only on the photometric properties of surrounding and contrasting objects, but also on the apparent *distance in depth* of its visible surfaces. In particular, whether the disks in the example appear uniformly white (as in 8a right), uniformly black (as in 8a left), or neither (as in 8b left and right), depends on whether conditions obtain such that a white surface appears to be visible through the medium of a dark surface or vice versa. The general phenomenological point to take away from this discussion is thus that the way a thing looks with respect to brightness can no more be extricated from the thing’s perceived placement in 3D space than can the way the thing looks with respect to shape. In both cases, depth is perceptually basic to visual experience.

7. Conclusion: The Sense-Datum Theory Refuted

In his recent ‘Real Presence,’ Noë chides philosophers who deny that the content of visual experience is ‘two-dimensional,’ i.e., that visual experience represents *both* objective spatial properties and apparent or ‘perspectival’ spatial properties:

Philosophers have tended to chicken out when it comes to the two-dimensional character of perceptual experience. They deny the legitimacy of one or the other dimensions of content. Sense-datum theorists deny that we visually experience the [tilted] coin’s circularity.... Direct realists, in contrast, may be tempted to deny that there is any sense in which it is *the coin* that looks elliptical.... (*forthcoming*, ms. 2).

In denying that when we see a 3D object we are able to see its P-shape, am I denying that there is an important sense in which the content of visual experience has a ‘perspectival’ dimension? Am I, as Noë puts it, ‘chickening out’?

The answer is: it depends. On the one hand, I reject, to quote A.D. Smith, ‘an account that postulates the existence of phenomenally 2D conscious states at a level that is prior to the operation of the [perceptual] constancies, and hence prior to 3D spatial awareness’ (2000b, 509). I reject the view that we are visually aware of 2D P-properties (and, hence, that P-properties are ‘perceptually basic’). P-properties are real, relation properties of visible objects, but, unless we are looking at a picture of an object, they are not properties that we can, in any literal sense, be said to see. Noë, like the classical sense-datum theorists, mistakes the act of *make-perceiving* an object’s P-shape for the act of seeing its P-shape (§4).

On the other hand, I do think that there is an important and perfectly

intuitive sense in which visual experience *does* have a perspectival dimension: When a perceiver views an object, what she sees is a function of the object's intrinsic spatial properties and her changing, extrinsic perspective on it. Indeed, if by an object's 'appearance,' we understand simply whatever 3D part or region of the object is visible from the perceiver's current point of view, then it is clear that the object's apparent spatial properties are perspectival. They vary systematically as the perceiver varies her perspective, and the perceiver is aware of this (Brewer 1997, chap. 6). The important point is that it is possible to allow that an object's apparent spatial properties are perspectival or 'egocentric' in this intuitive, low profile sense, but to deny that they are in any sense flat or 2D. Hence, contrary to Noë, there is no tension or 'apparent conflict' in visual experience between how an object perspectivally appears to the perceiver as she moves and her awareness of the object's actual, 3D spatial properties (2004, 78-79). An object's visual appearance or 'look' is simply the objective, environmentally situated, 3D part or region of the object that is visible from the perceiver's current point of view.

In *The Structure of Behavior*, Merleau-Ponty writes that 'when one speaks of the perspectival character of knowledge, the expression is equivocal':

It can signify that only the perspectival projection of objects would be given to primitive knowledge; and in this sense the expression is inexact since the first reactions of an infant are adapted, e.g., to the distance of objects – a fact that excludes the idea of a phenomenal world originally without depth.

At the same time, the perspectival character of knowledge, provides perception

...with the assurance of communicating with a world that is richer than what we know of it, that is, of communicating with a real world (1963, 186-87).

I am in full agreement with Merleau-Ponty that it is in the latter sense, but not the former, that seeing has an intrinsically perspectival dimension. This is not, as Noë puts it, to 'chicken out.' It is to reject a deeply rooted philosophical dogma about the distinction between appearance and reality in visual perception.

*Philosophy Department
Loyola University New Orleans*

REFERENCES

Ames, A. 1951: Visual perception and the rotating trapezoidal window. *Psychological Monographs*, 65 (7).

Andersen, R., L. Snyder, D. Bradley, and Zing, J. 1997: Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, 20, 303-330.

Anderson, B. 1999: Stereoscopic surface perception. *Neuron*, 24, 919-928.

—2003: The role of occlusion in the perception of depth, lightness, and opacity. *Psychological Review*, 110, 785-801.

Anderson, B. and Winawer, J. 2004: Image segmentation and lightness perception. *Nature*, 434, 79-83.

- Arterberry, M. and Yonas, A. 2000: Perception of 3D shape specified by optic flow by 8-week-old infants. *Perception and Psychophysics*, 62, 550–556.
- Atkinson, J., and Braddick, O. 1989: Development of basic visual functions. In A. Slater and G. Bremner (eds.), *Infant Development*. NJ: Erlbaum.
- Austin, J.L. 1962: *Sense and Sensibilia*. Oxford: Clarendon Press.
- Berkeley, G. 1709/1968: *A New Theory of Vision and Other Writings*. London: Dent.
- Bowyer, K. and Dyer, C. 1991: Aspect graphs: An introduction and survey of recent results. *International Journal of Imaging Systems and Technology*, 2, 315-328.
- Brewer, B. 1993: The integration of spatial vision and action. In N. Eilan, R. McCarthy, and B. Brewer (eds.), *Spatial Representation*. Oxford: Oxford University Press.
- 1997: *Perception and Reason*. Oxford: Oxford University Press.
- Briscoe, R. *forthcoming*: Egocentric spatial representation in action and perception. To appear in *Philosophy and Phenomenological Research*.
- Brooks, R. 1991: Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Bruce, V., P. Greene, and Georgeson, M. 2003: *Visual Perception: Physiology, Psychology and Ecology*, 4th edition. London: Psychology Press.
- Burge, T. 1986: Individualism and psychology. *The Philosophical Review*, 95, 3-45.
- Campbell, J. 1994: *Past, Space, and Self*. Cambridge, MA: MIT Press.
- Cheselden, W. 1728/1971: An account of some observations made by a young gentleman. In Pastore 1971, appendix A.
- Chisholm, R. 1957: *Perceiving: A Philosophical Study*. Ithaca: Cornell University Press.
- Clark, A. 1989: *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA: MIT Press.
- 1997: *Being There: Putting Brain, Body, and the World Together Again*. Cambridge, MA: MIT Press.
- 2001: Visual experience and motor action: Are the bonds too tight? *The Philosophical Review*, 110, 495-519.
- Colby, C. 1998: Action-oriented spatial reference frames in cortex. *Neuron*, 20, 15–24.
- Colby, C., J. Duhamel, and Goldberg, M. 1995: Oculocentric spatial representation in parietal cortex. *Cerebral Cortex*, 5, 470-481.
- Colby, C. and Goldberg, M. 1999: Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319–49.
- Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.
- 1985: Molyneux's question. In A. Phillips (ed.), *The Collected Papers of Gareth Evans* London: Oxford University Press.

- Farah, M. 2001: *Visual Agnosia*, 2nd ed. Cambridge, MA: MIT Press.
- Fleming, R. 2004: *Human Visual Perception Under Real-World Illumination*, unpublished PhD thesis, MIT Department of Brain and Cognitive Sciences.
- Fleming, R., O. Dror, and Adelson, E. 2003: Real-world illumination and the perception of surface reflectance properties. *Journal of Vision* 3, 347-368.
- Fleming, R., E. Torralba, and Adelson, E. 2004: Specular reflections and the perception of shape. *Journal of Vision*, 4, 798-820
- Fleming, R. and Anderson, B. 2004: The perceptual organization of depth. In L. Chalupa and J. Werner (eds.), *The Visual Neurosciences*. Cambridge, MA: MIT Press.
- Gallagher, S. 2005: *How the Body Shapes the Mind*. Oxford: Oxford University Press.
- Gazzaniga, M. 1998: *The Mind's Past*. Berkeley: The University of California Press.
- Gibson, J.J. 1966: *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.
- 1979: *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Gilchrist, A. 1977: Perceived lightness depends on perceived spatial arrangement. *Science*, 195, 185- 187.
- 1980: When does perceived lightness depend on spatial arrangement? *Perception and Psychophysics*, 28, 527-38.
- Gosselin, F. and Schyns, P. 2002: You are about to see pictorial representations! *Behavioral and Brain Sciences*, 25, 191-192.
- Graziano, M. and Botvinick, M. 2002: How the brain represents the body: Insights from neurophysiology and psychology. In W. Prinz and B. Hommel (eds.), *Common Mechanisms in Perception and Action: Attention and Performance XIX*. Oxford: Oxford University Press.
- Gregory, R. and Wallace, J. 1963: Recovery from early blindness: A case study. *Experimental Psychology Society, Monograph No. 2*.
- Grush, R. 1998: Skill and spatial content. In R. Grush (ed.), *The Philosophy of Gareth Evans. Electronic Journal of Analytic Philosophy* 6.
- 2000: Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind*, 1, 59–92.
- 2004: The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377-442.
- Helmholtz, H. von 1909/1925: *Physiological optics*, vol. 3. New York: Optical Society of America.
- Hoffman, D. 1998: *Visual Intelligence*. New York: W.W. Norton and Co.
- Howard, I. and Rogers, B. 1995: *Binocular Vision and Stereopsis*. Oxford: Oxford University Press.
- Hubel, D. and Wiesel, T. 1963: Receptive fields in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology*, 26, 994-1002.

- Hume, D. 1740/1975: *A Treatise of Human Understanding*, 2nd edition. Oxford: Clarendon Press.
- Hurley, S. and Noë, A. *forthcoming*: Can hunter-gatherers hear color? To appear in G. Brennan, R. Goodin, and M. Smith (eds.), *Common Minds: Essays in Honor of Philip Pettit*. Oxford: Oxford University Press.
- Jacob, P. and Jeannerod, M. 2003: *Ways of Seeing*. Oxford: Oxford University Press.
- Johnson, J. and Olshausen, B. 2005: The recognition of partially visible objects in the presence and absence of their occluders. *Vision Research*, 45, 3262-3276.
- Kanizsa, G. 1979: *Organization in Vision: Essays on Gestalt Psychology*. New York: Praeger.
- 1985: Seeing and thinking. *Acta Psychologica*, 59, 23-33.
- Kelly, S. 2004: On seeing things in Merleau-Ponty. In T. Carmon (ed.), *The Cambridge Companion to Merleau-Ponty*. Cambridge: Cambridge University Press.
- King, M., G. Meyer, J. Tangney, and Biederman, I. 1976: Shape constancy and a perceptual bias toward symmetry. *Perception and Psychophysics*, 19, 129-136.
- Knill, D. and Kersten, D. 1991: Apparent surface curvature affects lightness perception. *Nature*, 351, 228-30.
- Koch, C. 2004: *The Quest for Consciousness*. Englewood: Roberts and Co.
- Koenderink, J. 1984: The internal representation of solid shape and visual exploration. In L. Spillmann and B. Wooten (eds.), *Sensory Experience, Adaptation, and Perception*. New Jersey: Lawrence Erlbaum Associates.
- Koenderink, J. and van Doorn, A. 1979: The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32, 211-216.
- Koffka, K. 1935: *Principles of Gestalt Psychology*. Cleveland: Harcourt, Brace and World.
- Kosslyn, S. 1994: *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. and Sussman, A. 1995: Roles of imagery in perception: Or, there is no such thing as immaculate perception. In M. Gazzaniga (ed.), *The Cognitive Neurosciences*. Cambridge, MA: MIT Press.
- Kriegman, D. and Ponce, J. 1990: Computing exact aspect graphs of curved objects: Solids of revolution. *International Journal of Computer Vision*, 5, 119 – 135.
- Locke, J. 1689/1975: *An Essay Concerning Human Understanding*. Oxford: Oxford University Press.
- Mandik, P. 2005: Action-oriented representation. In K. Akins and A. Brook (eds.), *Cognition and the Brain: The Philosophy and Neuroscience Movement*. Cambridge: Cambridge University Press.
- Marr, D. 1982: *Vision*. New York: W.H. Freeman and Sons.
- Merleau-Ponty, M. 1945/1962: *The Phenomenology of Perception*. C. Smith, transl. London: Routledge and Kegan Paul Ltd.
- 1963: *The Structure of Behavior*. A. Fisher, transl. Boston: Beacon Press.

- Merriam, E. and Colby, C. 2005: Active vision in parietal and extrastriate cortex. *The Neuroscientist*, 11, 484-493.
- Merriam, E., R. Genovese, and Colby, C. 2003: Spatial updating in human parietal cortex. *Neuron*, 39, 361-373.
- 2007: Remapping in human visual cortex. *Journal of Neurophysiology*, 97, 1738-1755.
- Millikan, R. 1996: Pushmi-pullyu representations. In L. May, M. Friedman, and A. Clark (eds.), *Minds and Morals*. Cambridge, MA: MIT Press.
- 2004: *Varieties of Meaning*. Cambridge, MA: MIT Press.
- 2006: Postulating perceptual representations in a way that actually supports Gibson's central insights. Lecture at the Boston Colloquium for Philosophy of Science, January 30, 2006.
- Milner, D. and Goodale, M. 1995/2006: *The Visual Brain in Action*, 2nd edition. Oxford: Oxford University Press.
- Nakayama, K., Z. He, and Shimojo, S. 1995: Visual surface representation: A critical link between lower-level and higher-level vision. In S. Kosslyn and D. Osherson (eds.), *Visual Cognition*. Cambridge, MA: MIT Press.
- Noë, A. 2002: On what we see. *Pacific Philosophical Quarterly*, 83, 57-80.
- 2004: *Action in Perception*. Cambridge, MA: MIT Press.
- forthcoming: Real presence. To appear in a special issue of *Philosophical Topics*.
- O'Regan, K. 1992: Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46, 461-488.
- O'Regan, K. and Noë, A. 2001: A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 5.
- Palmer, S. 1999: *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.
- Peacocke, C. 1992: *A Study of Concepts*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. 2002: Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25, 157-238.
- 2003: *Seeing and Visualizing: Its Not What You Think*. Cambridge, MA: MIT Press.
- Rizzolatti, G., L. Fadiga, L. Fogassi, and Gallese, V. 1997: The space around us. *Science* 277, 190-191.
- Rock, I. 1983: *The Logic of Perception*. Cambridge, MA: MIT Press.
- Ruskin, J. 1856/1971: *The Elements of Drawing*. New York: Dover.
- Sacks, O. 1995: *An Anthropologist on Mars*. New York: Knopf.
- 2006: Stereo Sue: Why two eyes are better than one. *The New Yorker*, 82, 18, 64-74.
- Schor, C. 2004: Stereopsis. In L. Chalupa and J. Werner (eds.), *The Visual Neurosciences*. Cambridge, MA: MIT Press.

- Schwitzgebel, E. *forthcoming*: Do things look flat?
- Sellars, W. 1968: *Science and Metaphysics*. London: Routledge and Kegan Paul.
- Shimshoni, I. and Ponce, J. 1997: Finite-resolution aspect graphs of polyhedral objects. *IEEE Transaction Pattern Analysis and Mechanics Intelligence*, 19, 315-327.
- Siewert, C. 1998: *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.
- 2006: Is the appearance of shape protean? Available online at <http://psyche.cs.monash.edu.au/>.
- Simons, D. and Rensink, R. 2005a: Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9, 16-20.
- 2005b: Change blindness, representations, and consciousness: Reply to Noë. *Trends in Cognitive Sciences*, 9, 219.
- Slater, A., and Morison, V. 1985: Shape constancy and slant perception at birth. *Perception*, 14, 337-344.
- Slater, A., D. Rose, and Morison, V. 1984: Newborn infants perception of similarities and differences between 2- and 3D stimuli. *British Journal of Developmental Psychology*, 2, 287-294.
- Slater, A., A. Mattock, and Brown, E. 1990: Size constancy at birth: Newborn infants responses to retinal and real size. *Journal of Experimental Child Psychology*, 49, 314-322.
- Smith, A. 2000a: *The Problem of Perception*. Cambridge, MA: Harvard University Press.
- 2000b: Space and sight. *Mind*, 109, 481-518.
- Strawson, P. 1979: Perception and its objects. In G. MacDonald (ed.), *Perception and Identity*. Ithaca: Cornell University Press.
- Takeichi, H., T. Watanabe, and Shimojo, S. 1992: Illusory occluding contours and surface formation by depth propagation. *Perception*, 21, 177-184.
- Takeuchi, T. and Matsuoka, H. 2002: Material recognition under artificial illuminations. *Proceedings of ECVP, Perception* 31 (supplement), 135.
- Tarr, M. and Bühlhoff, H. (eds.) 1999: *Object Recognition in Man, Monkey, and Machine*. Cambridge, MA: MIT Press.
- Tarr, M. and Kriegman, D. 2001: What defines a view? *Vision Research*, 41, 1981-2004.
- Ullman, S. 1996: *High-Level Vision: Object Recognition and Visual Cognition*. Cambridge, MA: MIT Press.
- Van Effeltherre, T. 1994: Aspect graphs for visual recognition of 3D objects. *Perception*, 23, 563-582.
- Wexler, M. 2004: Two distinctions concerning emulators. *Behavioral and Brain Sciences*, 27, 422.
- Wiener, O. and Raab, T. 2004: Computing the motor-sensor map. *Behavioral and Brain Sciences*, 27, 423-424.
- Wiesel, T. 1982: Postnatal development of the visual cortex and the influence of environment.

Nature, 299, 583-92.

Wollheim, R. 1980: *Art and its Objects*, 2nd ed. Cambridge: Cambridge University Press.

—1998: On pictorial representation. *Journal of Aesthetics and Art Criticism*, 56, 217-226.

Word Count: 17,904