The Structure of Visual Content

Gabriel Greenberg
University of California, Los Angeles, USA
gabrielgreenberg@gmail.com

Visual representations express distinctively visual content. Such content takes the form of a kind of space where objects and properties are assigned locations in relation to a viewpoint. Many have conceived of visual space as a metric three-dimensional volume, analogous to physical space. Yet this assumption, I argue, over-constrains visual content, excluding the ubiquitous phenomenon of indeterminate depth perception. In this paper, I propose that visual contents are *view spaces*: two-dimensional directional arrays of objects, properties and relations. View spaces prioritize visual direction as a core aspect of structure, while demoting depth to a variable feature like colour or shape. This proposal accommodates depth indeterminacy while preserving distinctive visual structure, and it aligns with the use of feature maps in vision science and computer vision. I will argue that this architectural differentiation of direction and depth is supported by a wide range of evidence from philosophy and psychology.

In the vast landscape of representational forms, visual representations stand out as a natural class. States of visual perception, mental imagery, and visual memory; pictorial artefacts like photographs, drawings and maps; the feature maps of computer vision models— all seem to express a common type of visual content. In visual content, the attribution of properties and relations to objects is governed by spatial and perspectival organizing principles. It is widely accepted that visual content of this kind determines conditions of accuracy. Yet it also seems that such content cannot be propositional, at least not if propositions are understood to have language-like structure. But if the structure of visual content is not propositional in this sense, what is it? Does visual content have a distinctive form? This is the question this paper aims to address.[1]

A natural starting point is to conceive of visual content as a kind of visual space in which objects and properties occupy locations, and where the dimensions of this space are defined in relation to a central viewpoint. Many have interpreted these ideas metrically, conceiving of

[1] Thus the focal question concerns the nature of visual content, not the representational vehicles that carry this content. As a result, the present discussion does not bear directly on the question of the 'format' of perception and mental imagery, which is typically defined, in part, in terms of the structure of the underlying representational vehicles. See, for example, Goodman (1968, pp. 225–32), Camp (2007, pp. 154–60), and Quilty-Dunn (2019, pp. 3–6).

Mind, Vol. 00 . 0 . Month 2025
https://doi.org/10.1093/mind/fzaf061

visual space as a three-dimensional metrical volume, analogous to physical space.[2] In Peacocke's (1992, ch. 3) theory of *scenario content*, for example, objects are situated in perceptual space at determinate depths and directions relative to an origin point within a polar coordinate system. Although this spatially extended model captures important insights about the structure of visual space, I will argue that it ultimately imposes too much structure. Its rigid metrical shape is at odds with the ubiquitous perception of indeterminate depth.

Motivated by these considerations, this paper proposes a revision of scenario content, introducing instead a more flexible conception of visual structure in the form of a *view space*. View spaces are two-dimensional directional arrays, where each direction corresponds to a line of sight. Objects, properties and relations are located within regions of this array. A distinctive feature of view spaces is the prioritization of visual *direction* as a core structural aspect of visual space, while visual *depth* is demoted to the status of a variably represented feature, akin to colour, texture, shape or motion. This reconception of visual space introduces a degree of structural flexibility which smoothly accommodates perceptual depth indeterminacy, while preserving a distinctively visual structure and cohering with computational and psychological perspectives on vision.

The plan for the paper is as follows. §1 reviews existing approaches to visual contents along with cases of spatial perception that raise challenges for them. The theory of view spaces is introduced in §2, as a revision of metric scenario content. §3 defends the theory's key architectural assumption, that direction is a structural aspect of visual content, while depth is a represented feature. §4 discusses the integration of view space theory with our current scientific understanding of visual processing. §5 revisits the distinction between propositional and visual representation.

1. Visual content

This section discusses prominent contemporary approaches to visual content. In §1.1, I present possible-worlds analyses and review the motivations for shifting to a more structured framework. §1.2 introduces spatial approaches to visual content, focusing on Peacocke's theory of scenario content as a paradigm. Finally, §1.3 highlights structural challenges posed by indeterminate depth perception.

[2] See, for example, Marr (1982, pp. 275–83), Peacocke (1992, ch. 3), and Matthen (2005, pp. 271–89); see Wagner (2012, chs. 2–3) and Galebach (2018, ch. 1) for critical reviews.

1.1 Accuracy conditions

Like linguistic contents, the contents of visual representations characterize the world as being a certain way, and thereby determine conditions of accuracy. In a possible worlds framework, this means that accuracy for a visual representation is defined relative to a world. It is natural to think that the accuracy of visual representations also depends on the spatial perspective or *viewpoint* within a world, relative to which they are evaluated. A given visual representation may be accurate at world w, but only from viewpoint v. For example, a picture might be accurate at a viewpoint v close to a red cube and far from a blue sphere, but inaccurate at v', close to the blue sphere but far from the red cube. In this spirit, we may think of the accuracy conditions of visual representations as sets of *viewpoint-centred worlds*— that is, as sets of pairs of worlds and viewpoints.

The *unstructured approach* to visual content simply identifies visual contents with visual accuracy conditions. All and only those viewpoint-centred worlds relative to which a given visual representation is accurate are included in its content. This is the counterpart of unstructured possible worlds approaches to linguistic content. Theories of this kind are already familiar in the perception literature, and have gained traction in the philosophy of depiction.[3]

Such a view has its virtues: it provides an elegant framework in which accuracy and inaccuracy, as well as relations of entailment and consistency, can be defined in straightforward, set-theoretic terms. In addition, by rendering contents in the universal lingua franca of possible worlds, the theory makes the contents of visual and linguistic representations commensurable, elucidating semantic relations among representations of different perceptual modalities, between perception and thought, and between language and depiction. Thus, for many purposes, the unstructured approach to visual content identifies an important level of abstraction.

Nevertheless, in so far as we are seeking out an account of the nature of visual content specifically, the unstructured account has certain built-in limitations of *explanatory adequacy*. Precisely because it paints with such a general and modality-independent brush, unstructured content cannot provide an account of what is *distinctive* of visual content. Visual contents exhibit a characteristic spatial cohesiveness and perspectival organization. But nothing in the unstructured theory anticipates or explains this fact.

[3] For unstructured approaches to perceptual content, see Chalmers (2006) and Brogaard (2011). For pictorial content, see Ross (1997), Blumson (2009), Greenberg (2011), Abusch (2020), and Greenberg (2021).

Arbitrary sets of centred worlds exhibit no necessary spatial coherence. Even the viewpoints that make up the 'centres' of each centred world may be completely disjointed in space from one another; so arbitrary sets of centred worlds have no perspectival organization.

These observations are sufficient to motivate the search for a more structured account of visual content, one that captures what is distinctive of visual content. Such an account should, at least, give some sense of the spatial connectedness and the perspectival organization of visual content. The theory of visual content taken up in the second half of this paper aims to fulfil this mandate. It is not my goal here to argue against unstructured visual content once and for all. Indeed, my own view is pluralistic, allowing multiple levels of content for visual representations, just as multiple levels of abstraction are appropriate for any complex system. My contention instead is that there are aspects of content which cannot be captured at the unstructured level, and that these should be recovered at a further, structured level of visual content.

Besides their inherent explanatory limitations, unstructured approaches to visual content also face issues of *descriptive adequacy*. In a variety of cases, visual contents seem to mark distinctions more fine-grained than are available with sets of centred possible worlds, so cannot be described within the possible worlds framework. For example, differences in perceptual orientation (such as percepts of *square* vs. *diamond*), without any other changes to the spatial environment, resist modelling in terms of possible worlds.[4] Further argument can be made for the perception of spatially impossible objects. Since the unstructured view builds contents from possible worlds, as a matter of course, it is at a loss to characterize impossible contents of perception.[5]

[4] Peacocke (1992, pp. 75–6) draws attention to examples of variations in the perception of shape orientation, which are clearly represented by the visual system and phenomenologically accessible but do not make a difference to accuracy conditions. As Macpherson (2006, pp. 98–101) observes, the distinction cannot be teased apart modally, because every metaphysically possible situation in which a regular diamond is present is also one in which a tilted square is present, and vice versa. Parallel points extend to percepts of grouping.

[5] A number of potential examples have been noted in the literature, including: (i) the waterfall illusion (Crane 1988; Siegel 2024); (ii) overlapping reflections (Matthen 2005); (iii) the twofold character of picture perception (Wollheim 1987; Gregory 1970, p. 22); and (iv) impossible figures, such as the Penrose triangle (Penrose and Penrose 1958; Huffman 1971; Peacocke 1992, p. 74; Schacter et al. 1991). There is, in addition, widespread evidence of spatial inconsistencies in our everyday perception of distance, angle and collinearity. See Koenderink, Doorn and Lappin (2000, pp. 69–71), Suppes (2002, pp. 282–382), Todd and Norman (2003, pp. 41–4), Meadows (2011), Wagner (2012, chs. 4–5), Masrour (2017, pp. 6–9), and Galebach (2018, pp. 26–9). But because these findings are generally based on the comparison of judgements across two or more fixations, it is unclear to what extent they bear on visual representation specifically, as opposed to post-visual scene representation.

Although these descriptive issues won't be my primary focus in what follows, they are an important motivation for including, within visual content, some atomistic and quasi-propositional elements, along with a more geometric spatial structure.

1.2 Metric visual content

Whereas unstructured theories of content reify the conditions under which a representation is accurate, *structured approaches* aim to capture the form of content itself. In the case of visual content, this structure may be conceived as a *visual space*: a collection of individuals, properties and relations organized in three-dimensional spatial relations around a viewpoint (Indow 1991, pp. 430–1; Hatfield 2003, pp. 369–75).

The most straightforward realization of this idea construes visual space as one of the familiar spaces of geometry; such a space is governed by a metric, a relational structure of distances that give it shape. In a *metric visual space*, every object and property lies at a determinate *distance* and *direction* from the viewpoint, and thus from each other.[6] Such spaces are only partial, in the sense that they do not contain fully occluded objects or objects outside of the visual field, but they are nevertheless fully committal with respect to metric properties of the objects they contain, including their size, shape, distance and direction. Metric visual spaces are the sorts of things from which one could build a physical model by placing a model of each represented object at a determinate distance and direction from a defined origin.

Spaces of this kind seem to figure in accounts of the visual system that describe it as generating representations that reconstruct a three-dimensional model of the external world given the retinal input. Something like this view is commonly presupposed in vision science, where the output of perceptual computation is thought to be mapped to a three-dimensional coordinate space. When it is assumed that object locations are represented by coordinates in a coordinate space, a metric conception of visual content is presupposed (Galebach 2018, p. 6).

[6] See Galebach (2018, pp. 7–26) for a critique of 'visuo-perceptual metric space' and a review of arguments for it. Note that, in discussing metric visual spaces, I make no special assumption about the *kind* of metric at work. Over the years, philosophers and scientists have variously suggested that visual space is Euclidean, spherical and hyperbolic. See, for example, Koenderink, Doorn and Lappin (2000, pp. 69–71), Suppes (2002, pp. 282–382), Todd and Norman (2003, pp. 41–4), and Wagner (2012, chs. 4–5).

This idea finds especially complete philosophical expression in Peacocke's (1992, ch. 3) theory of *scenario content*.[7] Peacocke understands perceptual contents in terms of *scenarios*, glossed as 'ways of filling out space'. A scenario is defined as coordinate space for which (i) an egocentric origin, axes, and polar coordinate system have been identified (1992, p. 62); and (ii) for every distance and direction from this origin, the presence or absence of a surface is specified, along with the surface's orientation, texture, colour, illumination, solidity, and motion (p. 63).[8] Officially, Peacocke defines scenario content as a *set* of scenarios, where the shift to sets is intended to capture variation in perceptual acuity (p. 63). For now, I'll assume the simpler identification of scenario contents with individual scenarios, but I reprise the issue in the next section.

Peacocke's account of scenarios does not, on its own, resolve the issues of descriptive adequacy raised in the previous section. To handle these, Peacocke introduces a second layer of perceptual content, made up of *protopropositions*, elementary structures that each contain an individual and a property, or a series of individuals and a relation (1992, p. 77), with the constraint that the individuals in question be inhabitants of the lower-level scenario (p. 79, p. 241 n. 11). The idea is that the properties involved in protopropositions are sufficiently fine-grained to distinguish among physically identical ways of 'filling out space'. Thus, differences in object orientation (*tilted square* vs. *untilted regular diamond*), as well as differences in perceptual grouping, are to be accounted for by variation at the level of protopropositions. Likewise, the fine-grainedness of protopropositions, as well as their independence from one another, is leveraged to capture impossible contents (p. 79).[9]

As for the basic challenge of explanatory adequacy that faces unstructured content, scenario content resolves this by making the spatial

[7] See Matthen (2005, pp. 271–89; 2014, pp. 266–79) for another exemplar of the metrical space theory. Galebach (2018, pp. 11–29, nn. 15, 16) surveys a range of metric theories of visual space from the last century.

[8] Although Peacocke's subject is perceptual content in general, I will focus on its application to visual content. In this context, the egocentric origin of the coordinate system in a scenario plays roughly the role I've ascribed to viewpoint above.

[9] The use of protopropositions to address impossible *spatial* contents sits somewhat awkwardly within the scenario content framework. The analysis requires that the same kinds of spatial properties— like depth, orientation or shape— which appear in the lower-level scenario must also appear in the higher-level protopropositions in order to represent incompatible attributes. This raises questions of when properties and relations are reduplicated at both levels, and to what extent they should be. These are not fatal problems for the theory, but pitfalls I hope to avoid in the positive proposal below.

organization imposed by an origin, axis, and polar coordinate system part of its essential structure. There is no problem with explaining why arbitrary collections of possible worlds, or arbitrary collections of propositions, fail to define visual contents, for only highly constrained subsets of these correspond to genuine scenarios. Nevertheless, I believe that scenario content is something of an over-correction, for it introduces *too much* structure, demanding rigidity in visual content where we should have flexibility.

1.3 Indeterminate depth

The ubiquitous phenomenon of indeterminate depth perception poses a central challenge for any metric approach to visual content, including that of scenario content. The clearest instance of the problem arises in the perception of merely relative depth representation.

Under ideal viewing conditions, depth perception is nearly *absolute*, especially for nearby objects— that is, the visual system represents precise distances between objects and the viewpoint (Cutting and Vishton 1995, pp. 73–5, 100–2; Landy et al. 1995, p. 391; Sedgwick 1986, §2.2.5). But more often than not, perception is partially indeterminate with respect to metric depth. For example, there is the perception of *approximate depth*, where you perceive an object as located within some range of distances from the viewpoint, but not at any particular one. And especially important for our argument is the perception of merely *relative depth*, where you perceive one object as behind another, but not by how much (Landy et al. 1995, p. 392; Sedgwick 2005, p. 135; Koenderink, Doorn and Wagemans 2011, p. 543).[10]

Indeterminate aspects of depth perception have been widely studied in psychology (Cutting and Vishton 1995; Landy et al. 1995; Palmer 1999, p. 204). But they are already phenomenologically vivid to the human observer in cases where the visual information necessary for absolute depth perception is unavailable. Consider the perceptual experience one would have when looking at the scene in Figure 1 from a stationary position. One might have nearly absolute depth perception of the grass beneath one's feet (G) or the nearby pine bough (E). But the mid-ground and background afford less precision. We can see that the mountains in the background (C) are further away than the mid-ground trees (F), but we have no sense of exactly how far. Likewise, for the clouds (D), whose

[10] A richer iteration of relative depth is relative ratio depth perception, where you represent the ratio of metric distances between objects and the viewpoint, but not their absolute distance; see Koenderink, Doorn and Wagemans (2011, pp. 544–6).

Fig. 1. A scenic view.

details we can make out distinctly, we have no specific perceptual estimate of their distance from the viewpoint.

Perceptual scientists have documented a wide range of depth cues exploited by the visual system (Sedgwick 1986; 1996, ch. 7; Cutting and Vishton 1995; Palmer 1999, ch. 5). Some, like accommodation and convergence, can be used to compute the absolute depth of nearby objects (Cutting and Vishton 1995, pp. 79–110; Hershenson 1998, pp. 29–45; Palmer 1999, p. 204; Sedgwick 2005, pp. 140–52). But other cues, like occlusion, relative size, and texture gradients, provide only relative depth information (Sedgwick 1986, pp. 4–42; 2005, pp. 131–40; Cutting and Vishton 1995, pp. 81–9; Hershenson 1998, pp. 87–97). For example, occlusion tells us that the mountains (C) are more distant than the mid-ground trees (F), but not by how much. In general, indeterminate depth perception arises when there is sufficient visual evidence to determine that one object is further away than another but insufficient information to determine how much further (Loomis et al. 1996; Sedgwick 2022).

To dramatize the point, consider a visual environment, like Figure 2 below, that has been stripped of most natural depth cues, including motion, binocular information, height above the horizon, texture gradients, and familiar size. The scene nevertheless supports robust relative depth judgements, which seem to be grounded largely

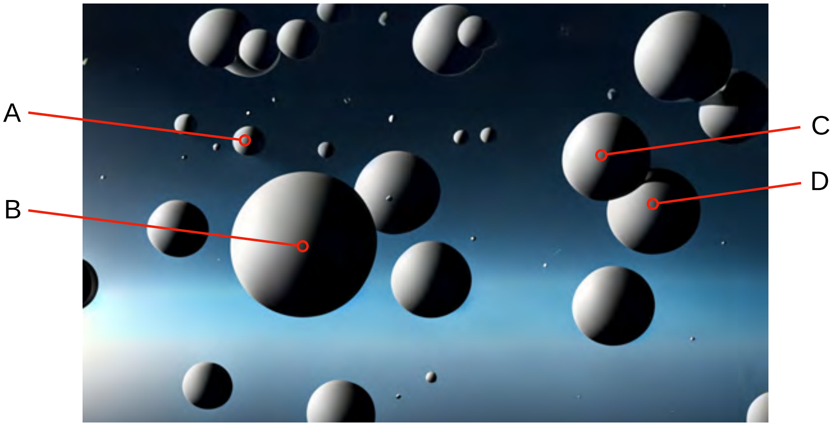

Fig. 2. An alien view. (Made with Adobe Firefly.)

in cues of occlusion and relative size. For example, relative size clearly indicates that orb A is more distant than orb B, while occlusion shows that orb D is more distant than orb C.

Indeterminacy with respect to depth typically implies indeterminacy about size as well, since if a given object (like a distant mountain) is closer, it must be smaller, or if further away, then larger (Sedgwick 1986, §2.2.6; Palmer 1999, pp. 315–27). Depth indeterminacy can also give rise to shape indeterminacy: if a perceiver does not determinately represent the depth relations between the several parts of a single object, she will not determinately represent its overall shape. For example, in the mid-ground tree (B) from Figure 1, because of the silhouetting effect, the relative depths of the various branches from the viewer are indeterminate, so the precise overall three-dimensional shape of the tree is indeterminate. Thus even relatively rich, mid-level visual perception seems to be capable of indeterminacy about the metric properties of depth, size and shape.

All of this raises a serious challenge for scenario content, or any account of visual space grounded in metric structure. The problem is that scenario content assigns determinate locations to all visible surfaces of all objects relative to a viewpoint. But indeterminacy in the perception of depth, size and shape undermines this metric specificity. They present perceptual features that could not be captured by a single physical model. The geometrical conception of visual space at the heart of scenario content must somehow be relaxed and revised.

One way to loosen the demands of scenario content is to allow that the locations of objects are given in content as *regions* of metric space, and objects are represented as located in the world *somewhere* within the indicated regions. This accommodation makes good sense of approximate depth perception, and plausibly models some cases of size indeterminacy. But even this revision cannot capture merely relative depth perception. Suppose the perceiver represents object *Y* as further than object *X* from the viewpoint. In this case, there is no boundary that can be drawn for *X* and *Y* respectively that could ensure the required ordinal relation while allowing it to be realized at the full range of distances from the viewpoint.[11]

A different way out of the problem is suggested in Peacocke's original formulation. Peacocke's official notion of *scenario content* is as a *set* of scenarios, rather than a single scenario (Peacocke 1992, p. 63). Peacocke introduces this flexibility in order to model variation in *perceptual acuity*, the degree of clarity or resolution in a perceptual state. But in principle, one could employ sets of scenarios to capture relative depth indeterminacy as well. If the content of a perceptual state was indeterminate with respect to relative depth, it would be modelled by a set of metrically determinate scenarios that differ only with respect to the depth relation in question.

Yet if scenario content is defined in terms of sets of scenarios, it thereby becomes an unstructured account of visual content, with all that that entails. As with sets of centred worlds, arbitrary sets of scenarios capture no spatial unity, central perspective point, or organizing spatial dimensions. But then scenario content alone is no longer in a position to explain these key aspects of visual content. In so far as the move from unstructured content to scenario content is motivated by the goal of providing a more explanatory account of visual content, then, under the set conception, some additional structure would be needed to specify which sets of scenarios correspond to visual content and which do not. And this structure evidently cannot take the strict metrical form of a scenario. But this just leads back to the question of what alternative structure is the right one.

In sum, the account of visual space as metric, exemplified by Peacocke's scenario content, can only achieve descriptive adequacy with respect to depth indeterminacy by giving up on its explanatory

[11] Consider two cases. First, suppose the *X*-locations are all nearer the viewpoint than the *Y*-locations; this accurately guarantees that *X* is closer than *Y*, but inaccurately excludes the possibility that *X* is located further away than anything in the *Y*-range. Second, suppose the *X*-locations and *Y*-locations overlap; then there is no guarantee that *X* is in fact closer than *Y*.

ambitions. This trade-off motivates the search for an alternative conception of visual structure that likewise captures the dimensions of fixity required by visual space but without imposing the over-restrictive structure of fully metric space.

2. View spaces

In this section, I outline a new account of the structural composition of visual content. It revises scenario content by holding fixed the *directional* structure of visual space, but relaxing its metrical structure in the dimension of *depth*; instead, depth is treated as a feature, like colour, shape or texture. This fact ultimately explains why depth can be indeterminate, like other visual features, and also why depth attributions can be spatially inconsistent. I refer to the resulting conception of visual contents as *view spaces*. View spaces are abstract structures populated with concrete objects, properties and relations.

View spaces are in many ways the visual counterparts of Russellian structured propositions (King 2007). Both are object-involving structures, and both determine precise accuracy conditions definable in terms of sets of centred worlds; in this sense, both are nominally 'propositional' (cf. Byrne 2001, pp. 201–2; Crane 2009; Grzankowski 2015; Camp 2018). Yet whereas structured propositions are fundamentally tree-like, made up of hierarchical binary branches, view spaces are basically array-like, their structure directional and geometrical. In this more specific sense, view spaces are not 'propositional' at all, but are a genuine alternative to structured content.

To this end, view spaces can be defined in two stages. First, there is an underlying spatial array, what I will call a *view field*, the content-level counterpart of phenomenology's visual field. Second, there are the objects, properties and relations which inhabit the view field.[12] In what follows, I introduce these elements, then proceed to the definition of accuracy for view spaces and the treatment of depth relations.

2.1 View fields

The basis of a view field is a two-dimensional surface of finite extent and continuous shape, whose specific form will vary among representational systems, as illustrated in Figure 3. We may call it the *view plane*.

[12] A comprehensive treatment would include the temporal extension of view spaces, reflecting the temporally extended contents of perception. Over time, view spaces could dynamically evolve, with constrained changes to the view field's shape and the distribution of objects and properties.

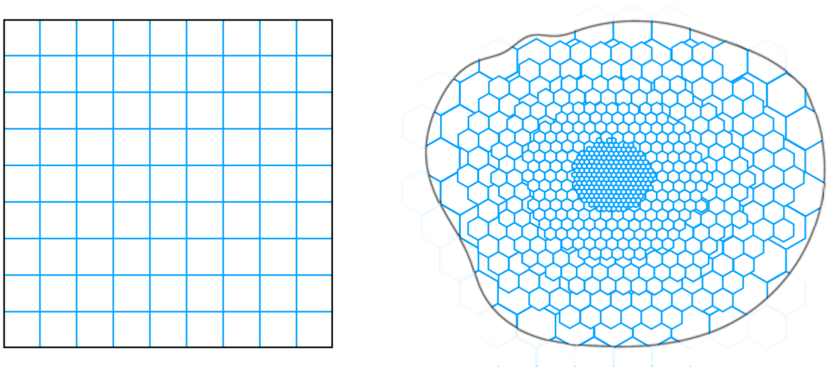

Fig. 3. *View planes.* On the left, a rectilinear view planewith a regular cell distribution, characteristic of digital pictures; on the right, an organic view plane with gradient cell density, styled after the human visual field.

(For some cases, this 'plane' may be curved.) In monocular human vision, for example, the view plane would form a kind of compressed half-oval, while in pictorial representation, it is normally flat and rectangular. It is divided exhaustively into potentially overlapping *cells*, which may, in principle, be as small as point-sized. In human vision, such cells would be more densely packed in the centre than the periphery, corresponding loosely to the greater focal acuity in the central regions of the visual field. For mechanical depiction, as in digital photography, the cells form a regular grid.

Every point on a view plane is associated with a *perspectival direction* that is oriented into the three-dimensional space surrounding the view plane. (Directions may be defined as rays whose endpoints are the points of the view plane.) The resulting structure is a *view field*. A view field defines a kind of directional space— a space whose 'dimensions', speaking loosely, are directions emanating from a view plane and whose extent is defined by the size and shape of the view plane. Thus the spatial anchor of view space is not the point-sized origin and axis of scenario content; rather, the counterpart of the 'viewpoint' is now a view field.[13]

[13] This definition of view field gives a precise rendering of Matthen's (2005, p.275) idea that 'visual directions constitute an omnipresent grid that overlays every scene, indexing the features represented in it'. Also compare Koenderink, Doorn and Wagemans (2011, p. 545): 'the pictorial space is a sheaf of depth threads'.

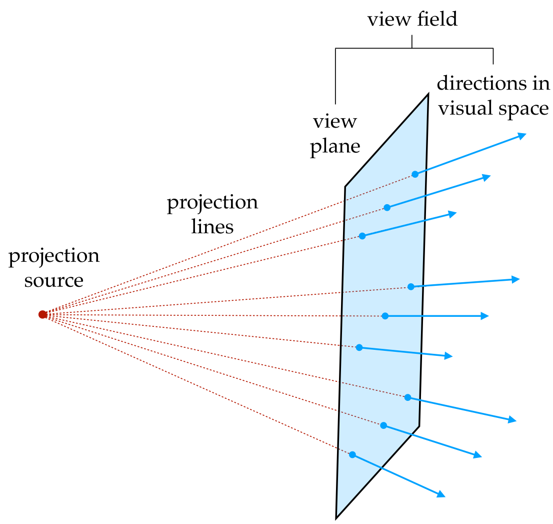

Fig. 4. View field with perspectival directions.

What makes the directions in a view field *perspectival* is that they have a kind of uniformity and coherence which is exemplified by linear perspective directions. In linear perspective, if we were to project each direction backwards through the view plane they would converge on a point, as in the traditional conception of a viewpoint. This is illustrated in Figure 4.

Describing the directional structure of the human visual field is subtle and beyond this paper's scope. In binocular vision the visual directions of each eye are fused in such a way that the perceptual origin point is the 'cyclopean eye' at the midpoint between the two eyes, resulting in an integrated set of visual directions (Hershenson 1998, pp. 14–27; Mapp, Ono and Howard 2012, pp. 230–48). I expect that complexities like this can be accommodated by reasonable extensions of the present account (Koenderink and Doorn 2008).

There are other kinds of view fields whose directions do not strictly converge backwards on a point, but still exemplify a perspective-like coherence. Indeed, it is unclear whether human vision itself follows such a strictly linear perspective (Helmholtz [1867] 1962, pp. 178–85, 328–30; Hansen 1973; Arnheim 1974; Hansen and Ward 1977; Rogers and Rogers 2009; Koenderink et al. 2010). Consider the examples from Figure 5: pictures in orthogonal projection can be interpreted

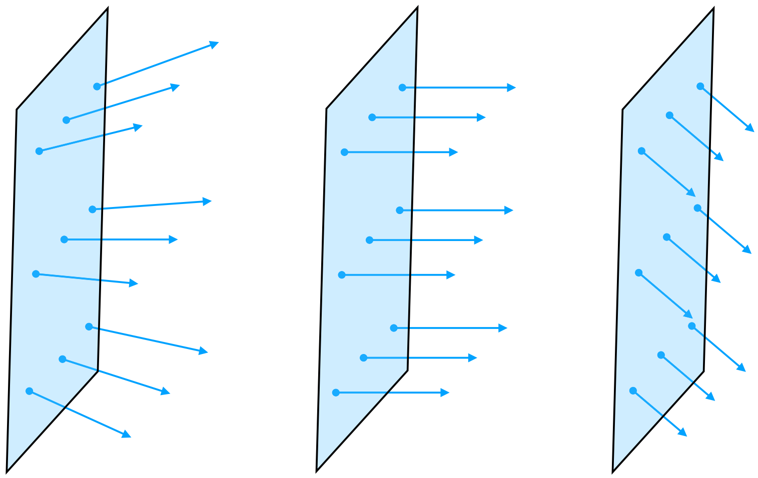

Fig. 5. View fields for different projective systems.

as expressing view spaces whose directions are all at right angles to the view plane, while those in oblique projection are all at the same oblique angle to the view plane. Curvilinear perspective, meanwhile, may be understood as involving view fields with curved view surfaces.[14]

Clearly, linear perspective, orthogonal, and oblique projections describe spaces in different ways. The present analysis locates this difference at the level of content in the overall directional shape of visual space. This is an aspect of view space content that is more fine-grained than accuracy conditions. Each type of image can display the same surfaces, in the same relative positions, plausibly expressing the same accuracy conditions, yet they differ in the overall directional organization of their spatial dimensions. Such a pictorial Frege-case calls for some kinds of 'modes of presentation', rendered here as directional arrays.

It is not clear how to capture these differences in scenario content, or any other model of content that takes the anchor of visual space to be an origin point. Orthogonal and oblique projections require a plane,

[14] My preferred approach to curvilinear perspective is to treat the the view 'plane' itself as curved; in that case, working out the spatial content of a curvilinear picture is partly a matter of mapping its featural content to this curved surface, before locating these features in directional space.

not a point, of projection. Meanwhile, curvilinear projections *are* taken from a point, so something else (like an intervening surface) must account for the intuitive difference in content between linear and curvilinear projections. These considerations offer further motivation for the idea that visual content is mediated by a two-dimensional planar surface, and not merely by directions originating at a viewpoint.

2.2 Feature clusters

A *view space* is a view field populated by objects, properties and relations. Objects and properties are collected together into *feature clusters*, defined as sequences containing a single object and a set of properties.[15] Each feature cluster is spatially anchored to the view field by association with a contiguous *cone of directions*: a set of directions that has a contiguous set of cells in the view plane as its footprint. Feature clusters may be associated with overlapping or nested cones; for example, if a complex object is associated with one cone, then each visible part of that object will typically be associated with a strictly nested cone.[16]

The objects in the feature clusters of a view space correspond to a representation's singular content, and the properties to part of its attributive content. (The remaining attributive content comes from the structure of the view space itself, along with relations, discussed below.) For present purposes, I treat feature clusters as including concrete individuals and properties as constituents, in the manner of Russellian propositions. But I believe a more fine-grained Fregean view is ultimately called for, where the singular elements are akin to senses or discourse referents (Burge 2005, pp. 6–9, 31–40; 2010, pp. 83–4, 380–1; Abusch 2013, pp. 16–17; Schellenberg 2018, pp. 84–101; Rescorla 2020, pp. 580–3).

Every feature cluster contains at least an object and a set of properties. For low-level visual representations, typical properties might include surface colours, illumination, motion, and attributes like *edge*

[15] The basic idea of a feature cluster is prefigured in much work on visual perception, especially beginning with Treisman's (1980) proposal that visual features are bound together into unified object representations. A number of philosophers have employed counterparts of the notion of object-property sequences, including 'feature clusters' (Pylyshyn 2003, p. 230), 'proto-propositions' (Peacocke 1992, p. 77), 'vectors of symbols' (Tye 2000, p. 91), 'feature-placing structures' (Matthen 2014, p. 272), 'multiple-slot memory' (Green and Quilty-Dunn 2021, pp. 678–87), and 'noun-phrase structures' (Burge 2018, pp. 90–1).

[16] Overlapping cones are a signature of amodal completion, since the cone of the completed object overlaps the cone of the occluder. In addition, more than one feature cluster can be associated with the same cone; representation of transparent surfaces that are perfectly aligned on the visual field will require this condition. See, for example, Pylyshyn (2003, pp. 192–3).

and *non-edge*. The corresponding objects would be low-level entities like small volumes, patches and surfaces. Higher-level representations may attribute properties of depth, shape, objecthood, and basic categories, and the corresponding objects would be relatively high-level entities reminiscent of cohesive mid-sized objects (Green 2018; Green and Quilty-Dunn 2021).

The inclusion of properties in feature clusters allows view spaces to make fine-grained distinctions, like that between attributions of *square* and *diamond*. Likewise, the potential for incompatible properties and relations associated with different feature clusters explains examples of impossible content. In this respect, they are similar to Peacocke's protopropositions.[17]

Ultimately, it may be necessary to expand the minimalist conception of feature clusters as sequences of objects and properties. A key question is whether there are object-specific aspects of structure that necessarily accompany the presentation of an object in visual content. Candidates include attributions of part–whole structure, volumetric shape, shape skeleton, object-centred coordinate frame, and topological or mereological structure (Palmer 1977; Marr and Nishihara 1978; L. Chen 2005; Feldman and Singh 2006; Green 2019; Lande 2020). I leave these questions to future work.

The final ingredients needed to define view spaces are relational features. Examples include relations between objects like *being the same size as* or *being a darker colour than*, but arguably the most important relations in visual representation are those of depth, discussed at greater length below. Unlike monadic features, relations cannot be structurally located *within* feature clusters, because feature clusters are associated with single objects, while relations hold between objects. Instead, I will conceive of relations as linking feature clusters together, as illustrated in Figure 6.

2.3 Accuracy

Taken as a whole, a view space locates each of the objects in its feature clusters in a given direction, and attributes to each its associated properties and relations. It is accurate when these attributions are correct. In effect, a view space displays its accuracy conditions across its surface.

[17] With one important caveat: in scenario content, protopopositions occupy a distinct level of content, and are not spatially tethered to the scenario content; Peacocke doesn't say explicitly how they are to be semantically linked (see Peacocke 1992, pp. 241–2 n. 11) In view spaces, feature clusters only enter the directional array by association with a region of this array, and this association carries the semantic attribution of visual direction.

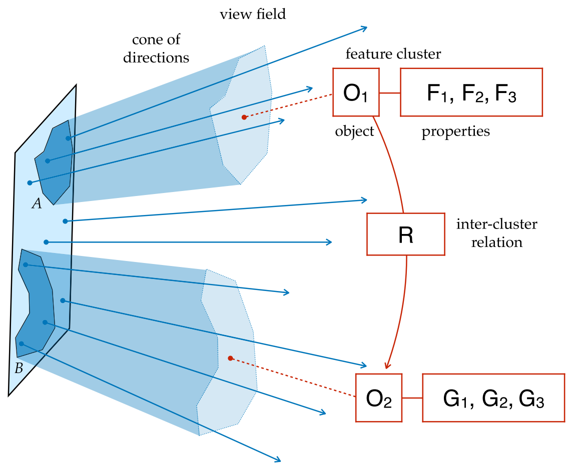

Fig. 6. A view space, including a view field with two feature clusters, and relation between them.

To define accuracy more precisely, we may adopt the idea that visual contents are accurate at viewpoint-centred worlds. Since, in the present framework, I have replaced viewpoints with view fields, I will say that a view space V is accurate relative to a world w and a positioned view field v_w, which is assigned a definite location and orientation within w (cf. Peacocke 1992, pp. 64–7. Then we can informally define accuracy as follows (cf. Lande forthcoming):

(1) A view space V is accurate at a world w and view field v_w if and only if, when V is embedded in w at v_w, for every feature cluster $F = \langle o, F_1, \ldots, F_n \rangle$ in V: (i) the object o is located within its associated cone of directions in w; (ii) o instantiates $F_1, \ldots, F_n$ in w; (iii) o stands in all of its associated relations to other objects in other feature clusters in w relative to v_w.

In addition, for an object to be *located within* a cone is for some part of it to be located in every cell of the cone (that is, the smallest cones which have view plane cells as their bases).[18] Variation in acuity across the visual field is the result of variation in the size of the corresponding cells. Where the cells are small, associated objects and properties are given

[18] One might experiment with different grades of the object-location relation for specific phenomena such as amodal completion or blur. A stronger relation could require that no part of the object be located outside the cone. A weaker definition would allow that that an object is located in a cone if some part of the object is located within the cone.

more precise directional locations; as the cells grow in size, the directional locations become less fine-grained.

2.4 Visual depth

View spaces depart from scenario contents, and any metrical conception of visual content, by their differential treatment of depth and direction. To occupy a position in a view space is to occupy a determinate direction relative to the view plane; by contrast, depth is understood as a relational feature, alongside colour, texture and shape, not part of the structure of visual content itself. In view spaces, only a kind of *primitive depth* is structural: all objects are located in the half-space defined by the directions emanating from the view plane. As a consequence, they are represented as somewhere 'out there', though how far 'out there' is not structurally defined.[19]

To capture more substantive attributions of depth, let us add to our semantic ontology relations between objects and the view plane. For example, an *absolute depth relation* is a binary relation of determinate distance between an object and the view plane; see Figure 7. Given that objects contain multiple points at different depths, I'll adopt the simplifying assumption that objects have unique *centre points*. Then, for a given embedding of a view field at a world:

(2) A relation R_m of absolute depth of magnitude m holds between view plane v and object o, $R_m(v, o)$ if and only if (i) there is a view field direction d that intersects v and the centre point of o, and (ii) v is distance m from the centre point of o along d.

Relative depth relations are relations between two or more objects and the view plane, as shown in Figure 7. Rather than specifying the absolute distance of an object from the view plane, such relations constrain the ordering in depth of the two objects relative to the view plane. We can understand relative depth in terms of relations that locate one object *between* the view plane and the other object.[20] Then we can define relative depth as follows.

[19] Koenderink, Doorn and Wagemans (2011, p. 543) suggests that this kind of depth perception is at work in the initial stage of viewing an undifferentiated Ganzfeld.

[20] The alternative is to think of relative depth in terms of comparative *further than* and *closer than* relations. In favour of the between-ness analysis is the fact that visual phenomenology does not seem to recognize a difference between *A* being more distant than *B* and *B* being closer than *A*. Thanks to Cian Dorr (p.c.) for noting this challenge, and to Sam Cumming (p.c.) for the resolution in the text. In principle, of course, comparative asymmetries might arise in nonconscious perceptual processing; see Codol (1990, pp. 395–6) for a possible example, and Galebach (2018, p. 27) for discussion.

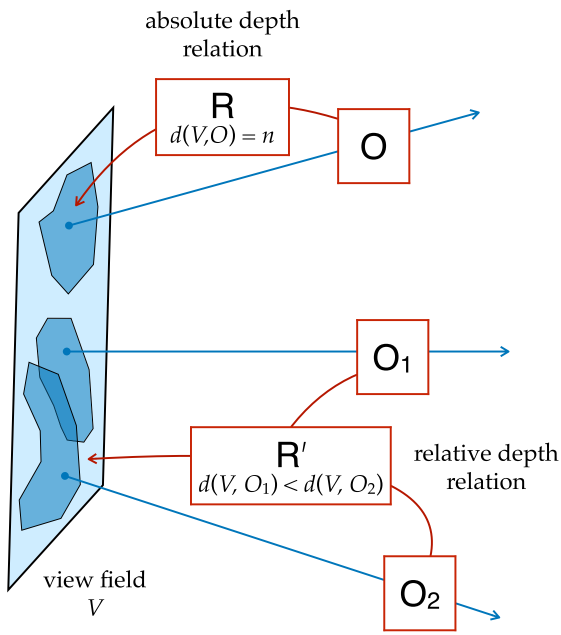

Fig. 7. A view space with absolute and relative depth relations.

(3) A relation R' of relative depth holds between the view plane v, and objects o_1 and o_2, $R'(v, o_1, o_2)$ if and only if (i) there is a view field direction d_1 that intersects v and the centre point of o_1, (ii) there is a view field direction d_2 that intersects v and the centre point of o_2, and (iii) the distance from v to the centre of o_1 along d_1 is less than the distance from v to the centre of o_2 along d_2.

In effect, this addition gives the answer to the problem of depth indeterminacy raised in the last section. View spaces can include depth information, while also allowing it to be metrically indeterminate, by including relative depth relations like the one defined above.

These ingredients also provide a plausible basis for those impossible visual contents that involve incompatible depth attributions, such as the Penrose triangle (Penrose and Penrose 1958). They correspond to view spaces that simultaneously include depth relations that are not mutually realizable in physical space.[21]

[21] Peacocke (1992, p. 79) likewise analyses the contents of 'impossible pictures' via the inclusion of incompatible properties at the level of protopropositions. This move may raise

3. Direction versus depth

In characterizing perceptual content, it is natural to make some demarcation between a relatively stable level of *structure* and the recombinable *features* which populate this structure. On such a scheme, the structure reflects the distinctive organization of the sensory modality in question— potentially differentiating between visual, auditory and somatosensory perceptual content— while the features correspond to the specific properties and relations attributed by a given representation on a given occasion.

It is in this spirit that Kant ([1781/1787] 1998, A24/B38, p. 175) describes the three-dimensional shape of perceptual space as a 'necessary representation' that provides 'the form of all appearances of outer sense' (A26/B42, p. 177). Likewise, Peacocke (1992, ch. 3) marks a distinction between the necessary ingredients of all scenarios— an origin, axes, and coordinate system— on the one hand, and the properties and protopropositions that populate them, on the other. The assumption in both cases is that the three-dimensional metric of visual space is part of its structure. This intellectual tradition effectively imports to perceptual space the intuitive conception of objective space, as a kind of three-dimensional metric container populated with objects and properties. While one can distinguish between direction and depth *coordinates* in such a space, this is just a way of describing unified three-dimensional locations.

The theory of view spaces breaks with this tradition by demoting attributions of depth to the status of feature, while retaining the organization of directions as structural.[22] Put another way, and speaking loosely, it holds on to the *X*- and *Y*-dimensions of three-dimensional metric space but renders distances in what would be the *Z*-dimension as relational features. We have already seen how these ideas make room in visual content for indeterminate depth attributions. In this section, I argue that four broadly psychological and computational considerations further support a basic bifurcation in the representational status of direction and depth.

methodological concerns of redundancy, since spatial relations like depth are now included in both the scenario and the protopropositions. It is unclear when depth is to be represented structurally and when as a feature. The view space account collapses the distinction.

[22] This kind of view is anticipated by Reid's distinction between 'original perception', exemplified by his two-dimensional 'geometry of visibles', and 'acquired perception', exemplified by perception of shape and depth (Van Cleve 2002; Copenhaver 2022). Thanks to Rebecca Copenhaver for this observation.

In drawing these connections between perceptual processing and content, I assume that a central theoretical role for content is to capture the informational properties of representational vehicles at a high level of abstraction. In the present case, my working hypothesis is that the structural aspects of visual content reflect architectural features of the computational system itself, which tend to be invariant and carry content in a way that is implicit. Meanwhile, featural aspects of visual content derive from representational elements of the computational system, which tend to be variable and explicit (cf. Hochman 2023). The proposal, illustrated in different ways below, is that view spaces build the semantic contribution of visual architecture into the directional structure of content, while the contribution of explicit depth representations is reflected in the featural constituents of content.

3.1 Distributional profiles

Initial evidence for the structural division between depth and direction comes from their distributional profiles, that is, how they vary within and across perceptual states (cf. Lande 2020, pp. 651–62). For example, Matthen (2005, pp. 274–6) highlights the following contrast: (a) in any visual scene, every visual direction (attributed within the visual field) is represented once and only once; whereas (b) it is not the case that in every visual scene, every depth relation (that is ever attributed) is represented once or only once— some depth attributes appear multiple times, in different directions, and many occur only in some but not other visual representations. Points parallel to (b) extend to colour, shape, texture and motion.[23]

From (a), Matthen concludes that 'the appearance of a visual direction is not empirically ascertained; it is a priori' (2005, p. 274). Apriority in turn is associated with structure. From (b), on the other hand, he remarks that 'distance is beginning to seem more and more analogous to colour, shape, or motion'. (Nevertheless, Matthen continues to make the metric space assumption that every represented surface is assigned a determinate depth.) He concludes:

> Visual directions constitute an omnipresent grid that overlays every scene, indexing the features represented in it. This is an updated

[23] Here is another contrast: visual contents always entail a *total* set of directional relations (as visual angles) between represented objects, but only a *partial* set of inter-object depth relations, since objects in different regions of the visual field may be incommensurable with respect to depth.

> version of Kant's argument about space: direction is part of the *form* of visual representation— this aspect of form arises from the feature maps of early vision— whereas features like red are part of the informative content. (Matthen 2005, p. 275)

Matthen's observation about the distribution of depths and directions seems to follow from a more general set of facts. To a first approximation, the distribution of directional attributes is a systematic and continuous function of visual field position, whereas depth attributes are not systematically related to visual field position at all, but vary across the visual field in ways that can only be predicted from the input to the visual system. In particular, the directions associated with the centre of the fovea point nearly straight ahead, gradually angling in more eccentric directions as they approach the periphery of the visual field, following the geometry of optical projection.[24] No such generalization relates depth to visual field position.

We can explain the fixed and systematic distribution of directional content by identifying it with the structure of visual content, and the variable distribution of depth content by identifying it with features in visual content. The dependence of direction on visual field position is rendered, at the level of content, as the correlation of directions with points on the view plane.

3.2 Computational profiles

A second set of differences between perception of depth and direction has to do with the computation and source of visual information.

In the first moments of vision, information about the wavelength of light reflected to the retina is registered by the variable activity of photoreceptor cells; but information about the visual direction of the source of that light is registered by the fixed *position* of each photoreceptor cell in the overall retinal layout.[25] Thus, from the most primitive point, information about visual direction is encoded in a way that is different from other kinds of visual information.

After the sensation of light, we can discern two important differences in the way that attributions of depth and visual direction are computationally derived. First there is the *kind* of computation involved. Vision can be thought of as implementing an extended solution to

[24] Note that attention seems to increase the resolution of directional cones in a given region of the visual field, but this does not distort the underlying distribution of unitary directions. See Carrasco (2011, pp. 1500–7).

[25] This point is anticipated by the nineteenth-century concept of perceptual 'local sign' (Lotze 1886, pp. 309–20; Koenderink and Doorn 2008, pp. 171–2).

the well-known inverse problem of reconstructing a 3-D scene from its 2-D retinal image. The computation of depth is a prime example of this, where depth attributions are the result of abductive and probabilistic computation, based on statistical 'cues' available in the retinal signal (Landy et al. 1995; Palmer 1999, ch. 5; Mertan, Duff and Unal 2022). Consequently, when depth cues are suppressed, errors in depth attribution increase markedly (Loomis et al. 1996). The same framework applies to nearly all properties and relations attributed by the perceptual system, including texture, motion, shape and colour.

An exception, however, is the representation of visual direction, which is largely determined by monotonic and context-insensitive inference from registered retinal location. For example, although perceived directions are the result of computationally merging distinct sets of ocular directions from each eye, this combination of information is generally achieved by fixed geometrical transformations and simple averaging (Mapp, Ono and Howard 2012). As a result, perceived visual direction is considered to be highly accurate, susceptible only to small errors induced by carefully constructed stimuli (Ono and Mapp 1995; Hershenson 1998, pp. 11–14; Mapp, Ono and Howard 2012, pp. 230–48). In so far as visual direction does require inference, these calculations are not aimed at reconstructing the 3-D world from a 2-D image, but at resolving physical ambiguities in the sense organs.[26]

Second, there is the *source* of the information processed in computation. The attribution of specific colours and shapes, as well as the attribution of depth, is worked out on the basis of visual cues derived from the incoming visual stream. The computation of such features depends in large part on information external to the visual system. By contrast, the assignment of visual directions, the shape of the view plane, and the resolution of the view field all depend primarily on internal aspects of the representing system. These include the fixed

[26] There are at least two areas where ampliative inference enters into the computation of perceived direction. First, the granularity of localization in the visual field outstrips the grain of retinal photoreceptors by interpolating intermediate visual locations, a phenomenon known as *hyperacuity* (Westheimer 1984, 2010; Smallman et al. 1996). Second, in the process by which directions from each eye are combined to determine a single set of *cyclopean* directions, the influence of each eye on the overall computation appears to be modulated by the quality of signal from each eye (Mansfield and Legge 1996; Mapp and Ono 1999). Still other findings suggest the subtle influence of visual motion, eye movement, attentional shift, and adaptation on perceived location (Bridgeman, Peery and Anand 1997; Whitaker, McGraw and Levi 1997; Ross et al. 2001; Schlag and Schlag-Rey 2002; Whitney 2002). Perhaps as a result of these effects, the potential for dissociation between encodings of retinal position and encodings of perceived location appear to grow modestly across the visual hierarchy (Fischer, Spotswood and Whitney 2011).

architecture of the visual system, along with variable inputs from the rotation and accommodation of the eyes, and locus of attention within the visual field (Hershenson 1998, pp. 11–14; Fischer, Spotswood and Whitney 2011; Mapp, Ono and Howard 2012, pp. 230–48.[27]

In sum, the computations that result in attributions of direction appear highly constrained and internally controlled, while those that result in attributions of depth are comparatively variable and externally dependent. We can see the structural status of direction and the featural status of depth as the content-level signature of these divergent underlying processes.

3.3 Psychological profiles

Recent work in perceptual science standardly draws a distinction between '2-D position' (or 'XY'-position) in the visual field and 'position-in-depth' ('Z'-position). We may suppose that each 2-D position corresponds to a narrow cone of directions in three-dimensional space. A wide range of psychological evidence indicates that depth and 2-D location are at least processed differently: not only are they sensitive to different cues and conditions, as we have seen, objects' 2-D locations are generally perceived faster and more accurately, and remembered more reliably than their locations in depth (Kasai et al. 2003; Umemura 2015; Cooper, Ginkel and Rokers 2016; Qian and Zhang 2019; Finlayson and Golomb 2017).

Moreover, in a series of papers, Golomb, Finlayson and colleagues have argued that there is something specifically structural about 2-D location, and featural about depth. Golomb, Kupitz and Thiemann (2014, pp. 2264–7, 2269–74) first showed that objects presented at the same 2-D location are more likely to be perceived as similar with respect to shape and colour, but neither of these attributes influences one another or location attribution in the same way. This result supports the widespread assumption that 2-D location plays a special, structural role in constituting object representations (Golomb, Kupitz and Thiemann 2014, pp. 2262–3). Importantly, Finlayson and Golomb (2016, 2017) went on to report that depth is likewise influenced by 2-D location, but not vice versa; and that depth does not influence colour, making it unlike 2-D location, but more like colour and shape. They conclude:

[27] This is true even for exogenously controlled attention; it is attention itself, not the interpretation of an input, that alters the acuity profile of the visual field. Even after the triggering input is gone, the shift in attention may persist.

> Our results suggest that despite the three-dimensional nature of our visual environment, only 2D location information— not position-in-depth— seems to be automatically bound to object features, with depth information processed more similarly to other features than to 2D location. (Finlayson and Golomb 2016, p. 49)

The broad outline of these claims is anticipated by a range of findings in neuroscience. For example, 2-D location is encoded throughout the visual system, including the growing number of early visual areas found to exhibit retinotopic organization (Grill-Spector and Malach 2004; Wandell, Dumoulin and Brewer 2007; Brewer and Barton 2012). Specialized sensitivity to depth, on the other hand, seems to be restricted to a handful of mid- and late-stage areas, such as V3A, V3B, V7, and MT (Neri, Bridge and Heeger 2004; Welchman et al. 2005, pp. 822–5; Preston et al. 2008; Finlayson, Zhang and Golomb 2017, p. 507; Berman 2018, pp. 83–99). So 2-D location and depth processing appear, at the outset, to have distinct neural underpinnings.

A study by Finlayson, Zhang and Golomb (2017) provides specific support for the segregation of depth and direction, through direct fMRI-based comparison. They found that while two-dimensional representations pervade the entire visual processing system, evidence for depth representation emerged only in later processing.[28] Importantly, their findings indicate that direction can be processed without depth, but depth is almost never processed without direction (Finlayson, Zhang and Golomb 2017, p. 515). This suggests that 2-D location is essential to visual processing in a way that depth is not.

Together, such evidence points to a pervasive architectural role for 2-D location in visual processing, and a more variable, representation-based role for depth attribution. The structure of view spaces reflects this fundamental division of semantic labour.

3.4 Representational format

The differential treatment of depth and direction is anticipated, finally, by the view that representations in visual perception are significantly *picture-like*. By this I mean at least (i) that the representational vehicle itself has the functional properties of a two-dimensional metrical

[28] Although disparity information is registered in V1, evidence suggests that these do not lead to stereoscopic depth representations until later in processing. See Cumming and Parker (1997); Preston et al. (2008); Barendregt et al. (2015). It remains unclear to what extent depth-related computations, such as amodal completion or robust figure-ground segmentation, occur in early vision. See, for example, Zhou, Friedman and von der Heydt (2000); Layton, Mingolla and Yazdanbakhsh (2012); Thielen et al. (2019).

surface; and (ii) that the metrical organization of this surface encodes information about the environment as a picture does, in part by functioning as an approximate perspective projection of visual space.

I don't wish to argue for this strong pictorialist hypothesis here, which remains controversial. Still, it is partially corroborated by recent arguments for the iconicity of representations in early vision, based on psychophysical evidence (Clarke 2021; Block 2023). An underlying picture-like representation is also anticipated by the widespread invocation of feature maps as representations in psychological accounts of early visual processing and in computer vision (Treisman 1986, 1988; Frisby and Stone 2010, ch. 10; Alzubaidi et al. 2021); see §4.2 for discussion. Finally, it is at least invited by observable facts of retinotopy, which reveal, all along the early and middle visual cortex, neural organization that recapitulates the flat, picture-like organization of the retina itself (Kosslyn, Thompson and Ganis 2006; Wandell, Dumoulin and Brewer 2007; Burge 2022; Buehler 2025).

In any case, it is clear that pictures themselves convey both visual direction and visual depth, but do so in radically different ways. This is brought out by the formal semantics of pictures, where the fundamental principle is that each picture is a geometrical projection of the visual space which it expresses as content (Schlenker 2018, pp. 402–4; Greenberg 2021; Patel-Grosz et al. 2023, pp. 641–2. Such a principle implies an association of each location on a picture's two- dimensional surface with a direction in three-dimensional visual space, irrespective of the kind of marking found at that location (Greenberg 2021, pp. 860–3). Thus a structural feature of syntax—metric location on a two-dimensional surface—is enlisted to encode direction, while other basic aspects of pictorial content, starting with the colours of surfaces, or the presence of edges, result from the interpretation of specific markings. More complex features, such as shape, category and depth, take both marking and mark configuration (distribution across 2-D locations) into account. So direction attribution is not only necessary, inherent in format, in a way that depth is not; it also flows from structural features of syntax, making an apt counterpart to structural features of content.

Pictorialism alone doesn't rule out the metric space hypothesis, because every object depicted could, in principle, be associated with a determinate depth. But it nevertheless sets up a structural contrast between depth and direction. By necessity, spatial positions in a picture-like representation represent directions. All further features, including depth, are dependent on the distribution of markers in a given representation.

These remarks about the semantic function of pictures carry over directly to informational encodings in the retina, which itself constitutes a two-dimensional surface of light registration. Here, direction information is structurally encoded but depth is not, and must be computed much later in the visual stream. Thus the retinal layout sets up an initial asymmetry between depth and direction at the earliest points in visual processing. Perceptual states that are downstream from the retina and preserve its spatial layout will always have direction information available to them, but only have depth if it is computed. In so far as later representations have a pictorial format, the asymmetry of direction and depth information would be sustained.[29]

4. Visual cognition

In this section, I highlight two ways the view space account sheds light on foundational questions about the nature of perceptual cognition: first, as a refinement to the doctrine of feature-placing; and second, as an account of how visual content evolves over the stages of perceptual processing.

4.1 Directional feature placing

Philosophers of perception have long thought that, for an object or property to be represented in perceptual experience, it must be assigned a location in perceptual space. This doctrine, known as *feature placing*, has a long history of interpretation and debate (Strawson 1959, pp. 202–4; Evans 1982, pp. 143–73; Peacocke 1992, pp. 71, 241–2 n. 11; Clark 2000, §2.6, ch.5; Pylyshyn 2007, pp. 91–8). The theory of view spaces gives a distinctive take on this issue, which we might call *directional feature placing*: for an object or property to be represented in a visual modality, it must be represented as located in a particular directional cone anchored at the viewpoint, but need not be represented in depth. In effect, feature placing is still required, but only when 'locations' are understood as directions, not as points or finite regions of three-dimensional space.[30]

[29] It would also introduce a complication into visual computations, since calculations of three-dimensional features, like depth, shape or size, would have to draw spatial information from both explicit representations and representational structure simultaneously.

[30] Another distinctive commitment of the view space theory is that features are always attributed to objects, not to locations themselves. Some (such as Clark) hold that the features are predicated of places; not so here. Features are predicates of objects (however primitive); but objects are always assigned a location. See Clark (2000, pp. 76–9, 164–6; 2004, pp. 447–53), Cohen (2004), and Matthen (2004, pp. 502–7).

The idea that visual perception of an object necessarily involves feature placing has its original support in phenomenology. In normal perception, we never experience a visual object except as located in a particular part of the visual field (hence located in a particular directional cone of visual space).

Such introspective evidence is complemented by empirical work on mid-level object representations known as *object files*, presumed responsible for our ability to track objects in our visual environment. A wide range of studies support the idea that object files are always encoded at specific 2-D locations in visual space, and cannot arise in the visual system without such locational encoding (Treisman and Gelade 1980; Tsal and Lavie 1993; Z. Chen 2009; Golomb, Kupitz and Thiemann 2014; Pertzov and Husain 2014; H. Chen and Wyble 2015).[31] Some research, like the work of Finlayson and Golomb (2016, 2017) discussed in §3.3, have further suggested that only 2-D location, and not 3-D position-in-depth, plays an essential role in locational binding. Collectively this research programme suggests that late-stage object files are bound to 2-D locations at the level of representation. View spaces provide the content-level counterpart of this generalization, in the form of directional feature placing.

Still, the question has sometimes arisen whether even 2-D locational binding is really necessary for feature representation. When objects are perceived outside of the focus of attention, featural mismatches become more likely— that is, features are seen in the wrong locations, typically grouped with the wrong objects. These results show that representations don't always keep objects and features in place. Some early commentators interpreted these findings to mean that it is possible to represent features without attributing any location to them at all, an apparent counterexample to feature placing.[32] But the emerging consensus is that featural mismatch doesn't imply that object features have no place, only that they are assigned to the wrong places; indeed, such false conjunctions are much more likely when two objects occupy nearby 2-D locations (Johnston and Pashler 1990; Ashby et al. 1996; Pashler 1999, pp. 97–9. Further, the fact that perception under taxing conditions results in *errors* of locational attribution, rather than non-specificity, indicates that directional feature placing remains the norm.

[31] See Finlayson and Golomb (2016, p. 49) for discussion and review of the literature.

[32] Treisman and Gelade (1980, p. 126), for example, conclude that such features may be 'free floating spatially'. See Galebach (2018, pp. 28–9) for a renewed defence.

A more foundational objection comes from Pylyshyn (2007, pp. 79–98), who rejects any kind of feature placing on conceptual grounds, and suggests that information about 2-D location is merely *registered*, not *represented*. Pylyshyn imagines a complex visual machinery that regularly exploits directional and locational information but never encodes it in content. Pylyshyn's austere picture of visual content is plausible if one focuses narrowly on the indexing and binding of object files, which are his primary concern. But it is less credible in light of the use that object files are put to by the rest of cognition. Object files are more than spatially indexed place-holders; they enable the localization of physical objects in the environment, the basis of all action in space (Golomb, Kupitz and Thiemann 2014, pp. 2262–4). The binding of object files to 2-D locations should be understood in spatial terms, as attributing perspectival directions to objects. Thus we still have good reason to accept the feature-placing implications of the view space approach.

4.2 Stages of visual processing

In the course of visual processing, there does not seem to be a single, distinguished representation which expresses the unique content of perception once and for all. Instead, there are many visual representations at different stages of perceptual processing, with different kinds of contents. Thus it is standardly thought that, in its early stages, the human vision system uses dense feature maps that tile the visual field with low-level features, while in later stages it maintains sparser representations that bind together features and categories into persisting object files. View spaces provide a common template for understanding visual content in these very different representational settings.

In early vision, *feature maps* are normally conceived as two-dimensional arrays, in which each cell contains a distinct symbol or numeral; collectively, these symbols register low-level features uniformly across the visual field. Distinct feature maps are posited for the detection of different feature dimensions, such as shape, colour, orientation, boundedness or motion. Feature maps are invoked variously as structural representations at the level of visual computation, physical structures at the level of neural implementation, and as data structures in computer vision models.[33]

[33] For computer vision applications, see Fukushima (1980), LeCun et al. (1989), and Alzubaidi et al. (2021); for discussion in psychology, see Treisman and Gelade (1980), Treisman (1986, 1988), and Frisby and Stone (2010, ch. 10).

Every feature map can be thought of as directly expressing a view space. The overall shape and layout of the feature map specifies the shape of the view plane, while the locations of the feature symbols give the locations of the corresponding feature clusters as projected onto the view plane.[34] What is left implicit is the relationship between 2-D positions on the feature map and directional cones in the view space. This is worked out by reverse projection, essentially like the case of picture-like representations discussed in §3.4, save that symbols, not markers, occupy the base positions (Greenberg 2019). Each cell corresponds to the base of a directional cone, or line of sight, projecting out from an implied viewpoint (Marr 1982, p. 283; Tye 2000, p. 81).

The view spaces expressed by feature maps exhibit four distinctive characteristics: (i) each feature cluster covers a very small region of the view plane; (ii) these feature clusters tile the entire view plane; (iii) all feature clusters contain the same type of property or feature dimension; and (iv) the feature clusters are minimal, with each feature cluster typically containing only a single feature. As we shall see, all of these characteristics are subject to revision.

As we move towards mid-level vision, perception turns increasingly towards larger and more complex units of structural organization, often interpolating forms that leave no direct trace on the retinal image. For example, in amodal completion, an occluding object blocks a part of an occluded object from view, yet viewers perceptually fill in the occluded boundaries (Singh 2004; Lande forthcoming). How can this phenomenon be captured within a view space, where each object is anchored to a 2-D region of the view plane? The answer is simply for represented objects to overlap their locations in directional space, even though they are also associated with different depth features, as shown in Figure 8. At the structural level of the view space, the occluded boundary has the same semantic status as the occluding boundary. Featural depth attribution ultimately distinguishes the locations of these objects in perceptual accuracy conditions.

In later ventral stream processing, so-called *object files* track the locations of mid-sized objects. Object files collect together the various features of persisting objects into unified and accessible memory registers.

[34] The singular representation of objects does not play an explicit role in the scientific theories that invoke feature maps, such as those of Treisman (1986); yet the view space account does require that every feature cluster contains a singular element, so some extension of the scientific account is required. Here, relatively small entities, like edges, patches, or parts of surfaces would typically instantiate the corresponding low-level features represented by feature maps, and these would form the basis of the feature clusters in view space.

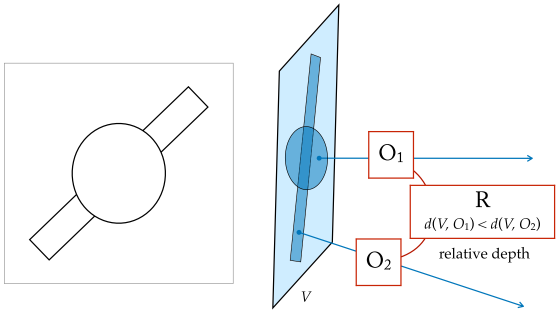

Fig. 8. An image that triggers amodal completion (left) and the corresponding view space with overlapping directional locations (right).

A prominent idea is that object files are the result of comparing and collating low-level feature maps, and identifying those clusters of detected features in the visual field likely to correspond to larger objects (Treisman 1986; Pylyshyn 2007; Green and Quilty-Dunn 2021). As a whole, the object file system is thought to be highly restricted, maintaining only three or four object representations at a time (Quilty-Dunn 2019, pp. 815–29; Green and Quilty-Dunn 2021, pp. 666–9).

Research on visual object representations has tended to focus on the internal characteristics of the memory structures involved, rather than their contribution to the broader representation of scene geometry. Still, it is widely recognized that the construction of object files essentially involves localization within the two-dimensional visual field, as discussed in §4.1. While little is yet known of the underlying representation format here, it is natural to interpret evidence for location binding as suggesting representations which fix a small number of object files within a larger visual field. This, in turn, is easily characterized at the level of content as a view space.

View spaces for object files will naturally locate the contents of each object file within a two-dimensional directional array. As only a small number of object files are ever maintained at once, the resulting view space will contain only a small number of total feature clusters. Each feature cluster will typically be associated with a relatively large region of the view plane, in comparison with the cells of low-level feature maps, but these will still fall far short of tiling the entire visual field. So gaps in the view field are inevitable. Even as feature clusters are sparser, they are also much richer, integrating together many different features at each condensation point. Their singular contents will be something in the family of bounded and cohesive bodies that figure in object cognition (Spelke 1990, pp. 48–50; Green 2018, pp. 179–80). On this scheme, each object file contributes a single rich feature cluster to the view space, via its association with regions of the view plane.

Parallel conceptions of sparse maps have emerged in models of visual imagery. Since imagery perforce has its origin in higher cognition, it is reasonable to expect that many aspects of imagery representations share basic properties with late-stage representations of object recognition and object files. Supporting this idea is the observation that imagery exhibits global visual organization but eschews dense 'photographic' detail (Block 1983, pp. 653–8). In Dennett's (1969, pp. 132–46) famous example, one is able to visually imagine the tiger, but not to count its stripes. It is plausible that the contents of mental images also exhibit the organization of sparse but rich visual fields that I have attributed here to object representations.[35]

A complementary perspective on late-stage perception, especially in the ventral stream, is that its central function is object recognition— both categorization and unique identification. These processes are thought to take place especially in the inferotemporal cortex (IT) in macaques, and homologous areas in the human lateral occipital cortex (DiCarlo, Zoccolan and Rust 2012). (I'll use 'IT' to refer to both.) IT serves as the main way-point for all processing in the ventral stream, so its representations are reasonable candidates for (one of the) 'final' outputs of the visual system. It is widely recognized that the retinotopic layout which characterizes earlier layers of vision is largely absent in IT (Grill-Spector and Malach 2004). This might lead to the expectation that the representations of IT, and thus the visual representations associated with object recognition, are divorced from the detailed spatial mapping that is presupposed by the view space account of visual content.

However, it turns out that IT is highly responsive to 2-D object position (DiCarlo and Maunsell 2003; Hong et al. 2016). It appears that, even as object recognition and identification are fine-tuned in IT, so are a wide variety of other, more spatialized properties, including size, pose, and 2-D position (Hong et al. 2016). We can provisionally conclude that the IT representations underlying object recognition are still strongly compatible with the feature placing hypothesis of the view space theory. To be sure, there are different representational strategies

[35] See, for example, Tye (2000, ch. 5) for an analysis of mental imagery in terms of sparse feature maps. Work on visual short term memory likewise suggests gappy scene representations (for example, Potter 1999; Potter et al. 2014).

employed in late- and early-stage vision, so the underlying mechanisms which ensure directional feature placing must undergo considerable evolution. Still, to my knowledge, the various representations within the ventral and dorsal pathways all support the same basic view of visual content.

In sum, we've seen how view spaces might take very different forms at different stages of visual representation. The view spaces of early visual representation are densely packed with uniform and shallow feature clusters, while those of late ventral stream representations appear to contain sparse arrangements of variable and deep feature clusters. Nonetheless, both situate their subject matters within essentially comparable arrays of viewpoint-centred directions; it is this commonality that the framework of view spaces brings to the fore.

My hypothesis is that all stages of visual perception express view spaces as content, and this is true for both ventral and dorsal streams. By contrast, other kinds of spatial representations, like the allocentric maps associated with the hippocampus, do not express view spaces, though they may contain view spaces as component parts (Tversky 2005; Greenberg 2025). Such conjectures must await further inquiry.

5. Mark of the visual

Using the tools developed in this essay, we may identify a central mark of the visual: to be a visual representation is to express content that takes the form of a view space. View spaces provide a well-defined alternative to the propositional conception of content rooted in the structure of language. We've seen how human visual perception fits this mould. Classifying visual representations by their contents makes it possible to abstract away from the cognitive function of perception, to include visual imagery, visual memory, and pictorial interpretation (Shepard and Podgorny 1978). As a result, the category of visual representation crosses the border from perception into cognition, by recognizing the commonality of vision and cognitively controlled visual imagery.

Identifying visual representations by their contents also suggests a degree of autonomy from their underlying representational format. While pictures clearly express view spaces, visual contents could also be described algebraically, using complex representations in a suitably rich formal language.[36] Or they could be encoded as the output vector

[36] Such representations should be distinguished from other kinds of computer graphics, which merely specify a two-dimensional image to be displayed on a monitor. While such *displays* are a type of picture, the underlying code merely represents the picture itself, not the picture's content.

of a computational vision model. In these cases, we would say that the linguistic description and the vector embedding may be classed as visual representations precisely because their contents specify view spaces. On this construal, the distinction between visual and non-visual representation cross-cuts the distinction between iconic and symbolic representation.[37]

In fact, the format-independence of visual representation is a key component of the view that perception itself has view space content. In the passage from early to late stages of visual perception, representation goes through a dramatic change in format, even if the nature of this evolution is not well understood. A corresponding evolution may be seen in convolutional neural network models of vision, from highly picture-like visual representations to abstract value vectors. We may say, in both cases, that these systems produce visual representations across their processing arcs, even as representational format changes at each layer.

With their distinctive directional organization, view spaces capture both the three-dimensional aspects of visual content and their origins in the two-dimensional layout of the retina. By preserving the structural signature of optical sensation while transcending its generative mechanics, view spaces reveal the deep continuity between perception and the diverse visual forms that bear its imprint.[38]

Conflict of Interest statement

None.

References

Abusch, Dorit 2013: 'Applying Discourse Semantics and Pragmatics to Co-reference in Picture Sequences'. *Proceedings of Sinn und Bedeutung*, 17, pp. 9–25.

[37] Compare Rescorla (2009) for a concept of map-like representation that is likewise independent of representational format.

[38] This paper has benefited from conversations with colleagues and students over many years, and from audiences at the Institut Jean-Nicod, University of Turin, London Aesthetics Forum, Eastern APA, Philosophy Mountain Workshop (with comments from Cian Dorr), Australian National University, Berkeley perception reading group, and the Toronto CPA. Thanks to Bill Kowalsky and Esther Ma for research and editorial assistance; to Jeff King for first posing to me the question this paper attempts to answer; and to Gail Ullman for providing the scenic view and a retreat for writing. Special appreciation to an anonymous reviewer whose knowledge, guidance and provocation led to the paper's current form.

———. 2020: 'Possible-Worlds Semantics for Pictures'. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, and Thomas Ede Zimmermann (eds.), *The Wiley Blackwell Companion to Semantics*, pp. 1–28. Hoboken, NJ: Wiley Blackwell.

Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Farhan Laith 2021: 'Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions'. *Journal of Big Data*, 8, pp. 1–74.

Arnheim, Rudolf 1974: 'Hansen's Curvilinear Perspective'. *Journal of Aesthetics and Art Criticism*, 32(3), p. 424.

Ashby, F. Gregory, William Prinzmetal, Richard Ivry, and W. Todd Maddox 1996: 'A Formal Theory of Feature Binding in Object Perception'. *Psychological Review*, 103(1), pp. 165–92.

Barendregt, Martijn, Ben M. Harvey, Bas Rokers, and Serge O, Dumoulin 2015: 'Transformation From a Retinal to a Cyclopean Representation in Human Visual Cortex'. *Current Biology*, 25(15), pp. 1982–7.

Berman, Daniel 2018: 'Representations of Spatial Frequency, Depth, and Higher-Level Image Content in Human Visual Cortex'. Ph.D. thesis, Ohio State University.

Block, Ned 1983: 'The Photographic Fallacy in the Debate About Mental Imagery'. *Noûs*, 17(4), pp. 651–61.

———. 2023: *The Border Between Seeing and Thinking*. Oxford: Oxford University Press.

Blumson, Ben 2009: 'Pictures, Perspective and Possibility'. *Philosophical Studies*, 149(2), pp. 135–51.

Brewer, Alyssa A., and Brian Barton 2012: 'Visual Field Map Organization in Human Visual Cortex'. In Stephane Molotchnikoff and Jean Rouat (eds.), *Visual Cortex: Current Status and Perspectives*, pp. 29–60. *Intech*. Available at http://dx.doi.org/10.5772/51914.

Bridgeman, Bruce, Shelley Peery, and Sulekha Anand 1997: 'Interaction of Cognitive and Sensorimotor Maps of Visual Space'. *Perception and Psychophysics*, 59(3), pp. 456–69.

Brogaard, Berit 2011: 'Centered Worlds and the Content of Perception'. In Steven D. Hales (ed.), *A Companion to Relativism*, pp. 137–58. Chichester: Wiley-Blackwell.

Buehler, Denis 2025: 'The Priority Map'. *Australasian Journal of Philosophy*, 103(1), pp. 235–60.

Burge, Tyler 2005: 'Disjunctivism and Perceptual Psychology'. *Philosophical Topics*, 33(1), pp. 1–78.

———. 2010: *Origins of Objectivity*. Oxford: Oxford University Press.
———. 2018: 'Iconic Representation: Maps, Pictures, and Perception'. In Shyam Wuppuluri and Francisco Antonio Doria (eds.), *The Map and the Territory: Exploring the Foundations of Science, Thought and Reality*, pp. 79–100. Berlin: Springer.
———. 2022: *Perception: First Form of Mind*. Oxford: Oxford University Press.
Byrne, Alex 2001: 'Intentionalism Defended'. *Philosophical Review*, 110(2), pp. 199–240.
Camp, Elisabeth 2007: 'Thinking With Maps'. *Philosophical Perspectives*, 21(1), pp. 145–82.
———. 2018: 'Why Maps Are Not Propositional'. In Alex Grzankowski and Michelle Montague (eds.), *Non-Propositional Intentionality*, pp. 19–45. Oxford: Oxford University Press.
Carrasco, Marisa 2011: 'Visual Attention: The Past 25 Years'. *Vision Research*, 51(13), pp. 1484–525.
Chalmers, David 2006: 'Perception and the Fall From Eden'. In Tamar Szabó Gendler and John Hawthorne (eds.), *Perceptual Experience*, pp. 49–125. Oxford: Oxford University Press.
Chen, Lin 2005: 'The Topological Approach to Perceptual Organization'. *Visual Cognition*, 12(4), pp. 553–637.
Chen, Zhe 2009: 'Not All Features Are Created Equal: Processing Asymmetries Between Location and Object Features'. *Vision Research*, 49(11), pp. 1481–91.
Chen, Hui, and Brad Wyble 2015: 'The Location but Not the Attributes of Visual Cues Are Automatically Encoded into Working Memory'. *Vision Research*, 107, pp. 76–85.
Clark, Austen 2000: *A Theory of Sentience*. Oxford: Clarendon Press.
———. 2004: 'Feature-Placing and Proto-Objects'. *Philosophical Psychology*, 17(4), pp. 443–69.
Clarke, Sam 2021: 'Mapping the Visual Icon'. *Philosophical Quarterly*. 72(3), pp. 552–77.
Codol, Jean-Paul 1990: 'Studies on Self-Centered Assimilation Processes'. In Jean-Paul Caverni, Jean-Marc Fabre, and Michel Gonzalez (eds.), *Cognitive Biases*, pp. 387–400. Amsterdam: North-Holland.
Cohen, Jonathan 2004: 'Objects, Places, and Perception'. *Philosophical Psychology*, 17(4), pp. 471–95.
Cooper, Emily A., Marcus van Ginkel, and Bas Rokers 2016: 'Sensitivity and Bias in the Discrimination of Two-Dimensional and Three-Dimensional Motion Direction'. *Journal of Vision*, 16(10), 5, pp. 1–11.

Copenhaver, Rebecca 2022: 'Berkeley and Reid'. In Samuel C. Rickless (ed.), *The Oxford Handbook of Berkeley*, pp. 615–36. Oxford: Oxford University Press.

Crane, Tim 1988: 'The Waterfall Illusion'. *Analysis*, 48(3), pp. 142–7.

———. 2009: 'Is Perception a Propositional Attitude?' *Philosophical Quarterly*, 59(236), pp. 452–69.

Cumming, Bruce G., and Andrew J. Parker 1997: 'Responses of Primary Visual Cortical Neurons to Binocular Disparity Without Depth Perception'. *Nature*, 389(6648), pp. 280–3.

Cutting, James E., and Peter M. Vishton 1995: 'Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information About Depth'. In William Epstein and Sheena Rogers (eds.), *Perception of Space and Motion*, pp. 69–117. San Diego: Academic Press.

Dennett, Daniel C. 1969: *Content and Consciousness*. London: Routledge and Kegan Paul.

DiCarlo, James J., and John H. R. Maunsell 2003: 'Anterior Inferotemporal Neurons of Monkeys Engaged in Object Recognition Can Be Highly Sensitive to Object Retinal Position'. *Journal of Neurophysiology*, 89(6), pp. 3264–78.

DiCarlo, James J., Davide Zoccolan, and Nicole C. Rust 2012: 'How Does the Brain Solve Visual Object Recognition?' *Neuron*, 73(3), pp. 415–34.

Evans, Gareth 1982: *The Varieties of Reference*. Edited by John McDowell. Oxford: Oxford University Press.

Feldman, Jacob, and Manish Singh 2006: 'Bayesian Estimation of the Shape Skeleton'. *Proceedings of the National Academy of Sciences*, 103(47), pp. 18014–19.

Finlayson, Nonie J., and Julie D. Golomb 2016: 'Feature-Location Binding in 3D: Feature Judgments Are Biased by 2D Location but Not Position-In-Depth'. *Vision Research*, 127, pp. 49–56.

———. 2017: '2D Location Biases Depth-From-Disparity Judgments but Not Vice Versa'. *Visual Cognition*, 25(9–10), pp. 841–52.

Finlayson, Nonie J., Xiaoli Zhang, and Julie D. Golomb 2017: 'Differential Patterns of 2D Location versus Depth Decoding Along the Visual Hierarchy'. *Neuroimage*, 147, pp. 507–16.

Fischer, Jason, Nicole Spotswood, and David Whitney 2011: 'The Emergence of Perceived Position in the Visual System'. *Journal of Cognitive Neuroscience*, 23(1), pp. 119–36.

Frisby, John P., and James V. Stone 2010: *Seeing: The Computational Approach to Biological Vision*. Cambridge, MA: MIT Press.

Fukushima, Kunihiko 1980: 'Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position'. *Biological Cybernetics*, 36(4), pp. 193–202.

Galebach, John Ethan 2018: 'Representing Geometry: Perception, Concepts, and Knowledge'. Ph.D. thesis, University of California, Irvine.

Golomb, Julie D., Colin N. Kupitz, and Carina T. Thiemann 2014: 'The Influence of Object Location on Identity: A Spatial Congruency Bias'. *Journal of Experimental Psychology: General*, 143(6), pp. 2262–78.

Goodman, Nelson 1968: *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: Bobbs-Merrill.

Green, E. J. 2018: 'What Do Object Files Pick Out?' *Philosophy of Science*, 85(2), pp. 177–200.

———. 2019: 'On the Perception of Structure'. *Noûs*, 53(3), pp. 564–92.

Green, E. J. and Jake Quilty-Dunn 2021: 'What Is an Object File?' *British Journal for the Philosophy of Science*, 72(3), pp. 665–99.

Greenberg, Gabriel 2011: 'The Semiotic Spectrum'. Ph.D. thesis, Rutgers University.

———. 2019: 'Tagging: Semantics at the Iconic/Symbolic Interface'. In Julian J. Schlöder, Dean McHugh, and Floris Roelofsen (eds.), *Proceedings of the 22nd Amsterdam Colloquium*, pp. 11–20. Amsterdam: University of Amsterdam.

———. 2021: 'Semantics of Pictorial Space'. *Review of Philosophy and Psychology*, 12(4), pp. 847–87.

———. 2025: 'Spatial Coherence in Narrative Film'. Unpublished manuscript, UCLA.

Gregory, R. L. 1970: *The Intelligent Eye*. New York: McGraw-Hill).

Grill-Spector, Kalanit, and Rafael Malach 2004: 'The Human Visual Cortex'. *Annual Review of Neuroscience*, 27, 649–77.

Grzankowski, Alex 2015: 'Pictures Have Propositional Content'. *Review of Philosophy and Psychology*, 6(1), pp. 151–63.

Hansen, Robert 1973: 'This Curving World: Hyperbolic Linear Perspective'. *Journal of Aesthetics and Art Criticism*, 32(2), pp. 147–61.

Hansen, Robert and John L. Ward 1977: 'Perception of Pictorial Space in Perspective Pictures'. *Leonardo*, 10(3), pp. 261–2.

Hatfield, Gary 2003: 'Representation and Constraints: The Inverse Problem and the Structure of Visual Space'. *Acta Psychologica*, 114(3), pp. 355–78.

Helmholtz, Hermann von [1867] 1962: *Treatise on Physiological Optics*. Edited by James P. C. Southall. New York: Dover Books.

Hershenson, Maurice 1998: 'Spatial Localization: Visual Directions'. In Maurice Hershenson (ed.), *Visual Space Perception: A Primer*, pp. 9–27. Cambridge, MA: MIT Press.

Hochman, Catherine 2023: 'Essential Indexicals in Perception'. Unpublished manuscript, UCLA.

Hong, Ha, Daniel L. K. Yamins, Najib J. Majaj, and James J. DiCarlo 2016: 'Explicit Information for Category-Orthogonal Object Properties Increases Along the Ventral Stream'. *Nature Neuroscience*, 19(4), pp. 613–22.

Huffman, David A. 1971: 'Impossible Objects as Nonsense Sentences'. *Machine Intelligence*, 6, pp. 295–323.

Indow, Tarow 1991: 'A Critical Review of Luneburg's Model with Regard to Global Structure of Visual Space'. *Psychological Review*, 98(3), pp. 430–53.

Johnston, James C., and Harold Pashler 1990: 'Close Binding of Identity and Location in Visual Feature Perception'. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), pp. 843–56.

Kant, Immanuel [1781/1787] 1998: *Critique of Pure Reason*. Edited and translated by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.

Kasai, Tetsuko, Takashi Morotomi Jun'ichi Katayama, and Takatsune Kumada (2003: 'Attending to a Location in Three-Dimensional Space Modulates Early ERPs'. *Cognitive Brain Research*, 17(2), pp. 273–85.

King, Jeffrey C. 2007: *The Nature and Structure of Content*. New York: Oxford University Press, USA).

Koenderink, Jan J., and Andrea J. van Doorn 2008: 'The Structure of Visual Spaces'. *Journal of Mathematical Imaging and Vision*, 31(2–3), pp. 171–87.

Koenderink, Jan J., Andrea J. van Doorn, Huib de Ridder, and Stijn Oomes 2010: 'Visual Rays Are Parallel'. *Perception*, 39(9), pp. 1163–71.

Koenderink, Jan J., Andrea J. van Doorn and Johan Wagemans 2011: 'Depth'. *I-Perception*, 2(6), pp. 541–64.

Koenderink, Jan J., Andrea J. van Doorn and Joseph S. Lappin 2000: 'Direct Measurement of the Curvature of Visual Space'. *Perception*, 29(1), pp. 69–79.

Kosslyn, Stephen M., William L. Thompson, and Giorgio Ganis 2006: *The Case for Mental Imagery*. Oxford: Oxford University Press.

Lande, Kevin J. 2020: ‘Mental Structures’. *Noûs*, 55(3), pp. 649–77.
———. forthcoming: ‘Contours of Vision: Towards a Compositional Semantics of Perception’. To appear in *British Journal for the Philosophy of Science*.
Landy, Michael S., Laurence T. Maloney, Elizabeth B. Johnston, and Mark Young 1995: ‘Measurement and Modeling of Depth Cue Combination: In Defense of Weak Fusion’. *Vision Research*, 35(3), pp. 389–412.
Layton, Oliver W., Ennio Mingolla, and Arash Yazdanbakhsh 2012: ‘Dynamic Coding of Border-Ow nership in Visual Cortex’ *Journal of Vision*, 12(13), 8, pp. 1–12.
LeCun, Yann, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel 1989: ‘Backpropagation Applied to Handwritten Zip Code Recognition’. *Neural Computation*, 1(4), pp. 541–51.
Loomis, Jack M., José A. Da Silva, John W. Philbeck, and Sergio S. Fukusima 1996: ‘Visual Perception of Location and Distance’. *Current Directions in Psychological Science*, 5(3), pp. 72–7.
Lotze, Hermann 1886: *Microcosmus*. New York: Scribner and Welford.
Macpherson, Fiona 2006: ‘Ambiguous Figures and the Content of Experience’. *Noûs*, 40(1), pp. 82–117.
Mansfield, J. Stephen, and Gordon E. Legge 1996: ‘The Binocular Computation of Visual Direction’. *Vision Research*, 36(1), pp. 27–41.
Mapp, Alistair P., and Hiroshi Ono 1999: ‘Wondering About the Wandering Cyclopean Eye’. *Vision Research*, 39(14), pp. 2381–6.
Mapp, Alistair P., Hiroshi Ono and Ian P. Howard 2012: ‘Binocular Visual Direction’. *Perceiving in Depth*, 2, pp. 230–48.
Marr, David 1982: *Vision*. New York: Henry Holt and Co.
Marr, David and Herbert Keith Nishihara 1978: ‘Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes’. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 200(1140), pp. 269–94.
Masrour, Farid 2017: ‘Space Perception, Visual Dissonance and the Fate of Standard Representationalism’. *Noûs*, 51(3), pp. 565–93.
Matthen, Mohan 2004: ‘Features, Places, and Things: Reflections on Austen Clark’s Theory of Sentience’. *Philosophical Psychology*, 17(4), pp. 497–518.
———. 2005: *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford: Clarendon Press.
———. 2014: ‘Image Content’. In Berit Brogaard (ed.), *Does Perception Have Content?*, pp. 265–90. New York: Oxford University Press.

Meadows, Phillip John 2011: 'Contemporary Arguments for a Geometry of Visual Experience'. *European Journal of Philosophy*, 19(3), pp. 408–30.

Mertan, Alican, Damien Jade Duff, and Gozde Unal 2022: 'Single Image Depth Estimation: An Overview'. *Digital Signal Processing*, 123(11), p. 103441.

Neri, Peter, Holly Bridge, and David J. Heeger 2004: 'Stereoscopic Processing of Absolute and Relative Disparity in Human Visual Cortex', *Journal of Neurophysiology*, 92(3), pp. 1880–91.

Ono, Hiroshi, and Alistair P. Mapp 1995: 'A Restatement and Modification of Wells-Hering's Laws of Visual Direction'. *Perception*, 24(2), pp. 237–52.

Palmer, Stephen E. 1977: 'Hierarchical Structure in Perceptual Representation'. *Cognitive Psychology*, 9(4), pp. 441–74.

———. 1999: *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

Pashler, Harold 1999: *The Psychology of Attention*. Cambridge, MA: MIT Press.

Patel-Grosz, Pritty, Salvador Mascarenhas, Emmanuel Chemla, and Philippe Schlenker 2023: 'Super Linguistics: An Introduction'. *Linguistics and Philosophy*, 46(4), pp. 627–92.

Peacocke, Christopher 1992: *A Study of Concepts*. Cambridge, MA: MIT Press.

Penrose, Lionel S., and Roger Penrose 1958: 'Impossible Objects: A Special Type of Visual Illusion'. *British Journal of Psychology*, 49(1), pp. 31–3.

Pertzov, Yoni, and Masud Husain 2014: 'The Privileged Role of Location in Visual Working Memory'. *Attention, Perception, and Psychophysics*, 76, pp. 1914–24.

Potter, Mary C. 1999: 'Understanding Sentences and Scenes: The Role of Conceptual Short-Term Memory'. In Veronika Coltheart (ed.), *Fleeting Memories: Cognition of Brief Visual Stimuli*, pp. 13–46. Cambridge, MA: MIT Press.

Potter, Mary C., Brad Wyble, Carl Erick Hagmann, and Emily S. McCourt 2014: 'Detecting Meaning in RSVP at 13 ms per Picture'. *Attention, Perception, and Psychophysics*, 76, pp. 270–9.

Preston, Tim J., Sheng Li, Zoe Kourtzi, and Andrew E. Welchman 2008: 'Multivoxel Pattern Selectivity for Perceptually Relevant Binocular Disparities in the Human Brain'. *Journal of Neuroscience*, 28(44), pp. 11315–27.

Pylyshyn, Zenon 2003: *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.
———. 2007: *Things and Places: How the Mind Connects with the World*. Cambridge, MA: MIT Press.
Qian, Jiehui, and Ke Zhang 2019: 'Working Memory for Stereoscopic Depth Is Limited and Imprecise— Evidence from a Change Detection Task'. *Psychonomic Bulletin and Review*, 26, pp. 1657–65.
Quilty-Dunn, Jake 2019: 'Perceptual Pluralism'. *Noûs*, 54(4), pp. 807–38.
Rescorla, Michael 2009: 'Cognitive Maps and the Language of Thought'. *British Journal for the Philosophy of Science*, 60(2), pp. 377–407.
———. 2020: 'Perceptual Co-Reference', *Review of Philosophy and Psychology*, 11(3), pp. 569–89.
Rogers, Brian, and Cassandra Rogers 2009: 'Visual Globes, Celestial Spheres, and the Perception of Straight and Parallel Lines'. *Perception*, 38(9), pp. 1295–312.
Ross, Jeff 1997: *The Semantics of Media*. Dordrecht: Kluwer Academic.
Ross, John, M. Concetta Morrone, Michael E. Goldberg, and David C. Burr 2001: 'Changes in Visual Perception at the Time of Saccades'. *Trends in Neurosciences*, 24(2), pp. 113–21.
Schacter, Daniel L., Lynn A. Cooper, Suzanne M. Delaney, Mary A. Peterson, and Mindy Tharan 1991: 'Implicit Memory for Possible and Impossible Objects: Constraints on the Construction of Structural Descriptions'. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), pp. 3.
Schellenberg, Susanna 2018: *The Unity of Perception: Content, Consciousness, Evidence*. Oxford: Oxford University Press.
Schlag, John, and Madeleine Schlag-Rey 2002: 'Through the Eye, Slowly: Delays and Localization Errors in the Visual System'. *Nature Reviews Neuroscience*, 3(3), p. 191.
Schlenker, Philippe 2018: 'What Is Super Semantics?' *Philosophical Perspectives*, 32(1), pp. 365–453.
Sedgwick, H. A. 1986: 'Space Perception'. In Kenneth R. Boff, Lloyd Kaufman, and James P. Thomas (eds.), *Handbook of Perception and Human Performance, Vol. 1: Sensory Processes and Perception*, ch. 21, pp. 1–57. Hoboken, NJ: Wiley.
———. 2005: 'Visual Space Perception'. In E. Bruce Goldstein (ed.), *Blackwell Handbook of Sensation and Perception*, pp. 128–67. Malden, MA and Oxford: Blackwell.
Sedgwick H. A. 2022: 'Ibn al-Haytham's Ground Theory of Distance Perception'. *i-Perception*, 13(5):1–17.

Shepard, Roger N., and Peter Podgorny 1978: 'Cognitive Processes that Resemble Perceptual Processes'. In William K. Estes (ed.), *Handbook of Learning and Cognitive Processes, V: Human Information Processing*, pp. 189–237. Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegel, Susanna 2024: 'The Contents of Perception'. In Edward N. Zalta and Uri Nodelman (eds.), *The Stanford Encyclopedia of Philosophy (Fall 2024 edition)*. https://plato.stanford.edu/archives/fall2024/entries/perception-contents/.

Singh, Manish 2004: 'Modal and Amodal Completion Generate Different Shapes'. *Psychological Science*, 15(7), pp. 454–9.

Smallman, H. S., D. I. MacLeod, S. He, and R. W. Kentridge 1996: 'Fine Grain of the Neural Representation of Human Spatial Vision'. *Journal of Neuroscience*, 16(5), pp. 1852–9.

Solso, Robert L. 1996: *Cognition and the Visual Arts*. Cambridge, MA: MIT Press.

Spelke, Elizabeth S. 1990: 'Principles of Object Perception'. *Cognitive Science*, 14(1), pp. 29–56.

Strawson, P. F. 1959: *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.

Suppes, Patrick 2002: *Representation and Invariance of Scientific Structures*. Stanford, CA: CSLI Publications.

Thielen, Jordy, Sander E. Bosch, Tessa M. van Leeuwen, Marcel A. J. van Gerven, and Rob van Lier 2019: 'Neuroimaging Findings on Amodal Completion: A Review'. *I-Perception*, 10(2), 2041669519840047.

Todd, James T., and J. Farley Norman 2003: 'The Visual Perception of 3-D Shape from Multiple Cues: Are Observers Capable of Perceiving Metric Structure?' *Perception and Psychophysics*, 65(1), pp. 31–47.

Treisman, Anne (1986: 'Features and Objects in Visual Processing'. *Scientific American*, 255(5), pp. 114B–25.

———. 1988: 'Features and Objects: The Fourteenth Bartlett Memorial Lecture'. *Quarterly Journal of Experimental Psychology*, 40(2), pp. 201–37.

Treisman, Anne and Garry Gelade 1980: 'A Feature-Integration Theory of Attention'. *Cognitive Psychology*, 12(1), pp. 97–136.

Tsal, Yehoshua, and Nilli Lavie 1993: 'Location Dominance in Attending to Color and Shape'. *Journal of Experimental Psychology: Human Perception and Performance*, 19(1), pp. 131–9.

Tversky, Barbara 2005: 'Cognitive Maps, Cognitive Collages, and Spatial Mental Models'. In Andrew U. Frank and Irene Campari (eds.), *Spatial Information Theory: A Theoretical Basis for GIS*, pp. 14–24. Berlin and Heidelberg: Springer.

Tye, Michael 2000: *The Imagery Debate*. Cambridge, MA: MIT Press.

Umemura, Hiroyuki 2015: ‘Independent Effects of 2-D and 3-D Locations of Stimuli in a 3-D Display on Response Speed in a Simon Task’. *Frontiers in Psychology*, 6, 1302.

Van Cleve, James 2002: ‘Thomas Reid’s Geometry of Visibles’. *Philosophical Review*, 111(3), pp. 373–416.

Wagner, Mark 2012: *The Geometries of Visual Space*. Hove: Psychology Press.

Wandell, Brian A., Serge O. Dumoulin, and Alyssa A. Brewer 2007: ‘Visual Field Maps in Human Cortex’. *Neuron*, 56(2), pp. 366–83.

Welchman, Andrew E., Arne Deubelius, Verena Conrad, Heinrich H. Bülthoff, and Zoe Kourtzi 2005: ‘3D Shape Perception from Combined Depth Cues in Human Visual Cortex’. *Nature Neuroscience*, 8(6), pp. 820–7.

Westheimer, Gerald 1984: ‘Spatial Vision’. *Annual Review of Psychology*, 35(1), pp. 201–26.

———. 2010: ‘Visual Acuity and Hyperacuity’. In Michael Bass (ed.), *Handbook of Optics, III: Vision and Vison Optics*. New York: McGraw Hill.

Whitaker, David, Paul V. McGraw, and Dennis M. Levi 1997: ‘The Influence of Adaptation on Perceived Visual Location’. *Vision Research*, 37(16), pp. 2207–16.

Whitney, David 2002: ‘The Influence of Visual Motion on Perceived Position’. *Trends in Cognitive Sciences*, 6(5), pp. 211–6.

Wollheim, Richard 1987: *Painting as an Art*. London: Thames and Hudson.

Zhou, Hong, Howard S. Friedman, and Rüdiger von der Heydt 2000: ‘Coding of Border Ownership in Monkey Visual Cortex’. *Journal of Neuroscience*, 20(17), pp. 6594–611.

