The Structure of Visual Content

Gabriel Greenberg   November 6, 2019

It is widely thought that visual representations express contents with accuracy conditions. But it also seems that these contents cannot be propositional, at least not if propositions are understood to have the kind of hierarchical syntactic structures typically associated with language (Crane 2009; Camp 2018). But if the structure of visual content is not propositional in this sense, what is it? What sorts of things are visual contents anyway? This is the question this essay attempts to answer.

I'll begin by considering, and arguing against, two prominent answers from the literature. One, represented by Peacocke's theory of scenario content, treats visual contents as three-dimensional spaces, filled out with arrangements of objects and properties. The other treats visual contents as sets of centered possible worlds. Both, I'll argue, face counter-examples from central phenomena in visual perception and depiction. In their stead, I'll defend a *structured* theory of visual content. Whereas structured propositions are thought to have language-like structures, I'll argue that visual contents take the form of *perspectival feature maps*, a kind of directional array populated with objects, properties, and relations. Building on the notion of a feature map from vision science, perspectival feature maps offer a new way forward in the analysis of visual content.

Section 1 defines the scope of a theory of visual content. In Section 2, I argue that theories which identify visual contents with concrete, spatial scenarios (e.g. Peacocke 1992) are unable to capture important cases of spatial indeterminacy. And in Section 3, I argue that unstructured theories of visual content, like those which identify perceptual content with sets of centered-worlds (e.g. Chalmers et al. 2006, Brogaard 2011) falter over cases of impossible images and contradictory percepts. Structured visual content addresses both concerns. Section 4 lays out the key commitments and motivations of the perspectival feature map approach, while 5 compares the proposal to other kinds of structured content. Finally, in Section 6, I discuss applications of the analysis to central phenomena in visual perception and mental imagery.

1 Visual content

Among the wide range of types of representation to be found in signaling systems and in the mind, VISUAL REPRESENTATION seems to be a natural class, and its members seem to express a common type of content, VISUAL CONTENT. The original case of visual content is that of visual perception, and this will be my primary case study in the discussion to follow. I assume a broadly representationalist approach of visual perception. That is, I assume that perception involves the tokening of perceptual representations, themselves the vehicles or bearers of visual content, and that these perceptual representations serve as inputs or outputs to computations carried out by the visual system. On this account, vision itself takes place in multiple stages, including conscious perceptual experience as well as various sub-personal representational states of the visual system. The conjecture pursued here is that both conscious visual perception and unconscious visual subsystems express a common kind of visual content.

Beyond perception, visual representation includes both mental imagery and visual memory. There are also artifactual visual representations, namely pictures— including photographs, drawings and paintings, and many maps. Certain kinds of data structures in computer vision should be considered visual representations as well. Despite the great diversity of these representations, all seem to express content of a characteristically visual type, in which objects and features are arranged in space around a central viewpoint or perspective.

In the essay I will develop the proposal that visual contents are *perspectival feature maps*, a notion I will explicate shortly. To be clear, my central claim is that visual *contents* have a feature-map structure, not that the representational *vehicles* which express visual content have this structure. Thus my proposal does not bear directly on the traditional question of the "format" of perception and mental imagery, which, at least on one reading, asks after the nature of representational vehicles. Instead, my claim is that visual representations, whatever their physical realization or vehicular structure, express a common type of content, and my aim here is to identify its essential features.

2 Visual content as metric space

In this section and the next, I examine two quite different but natural approaches to modeling visual content. The first, covered here, suggests that visual contents take the form of a type of space, a VISUAL SPACE, in which a three-dimensional array centered at a viewpoint or origin is populated with individuals, properties, and relations (e.g. Luneburg 1947; Suppes 1977; Rogers 1995; Koenderink and Doorn 2008; Wagner 2012; Erkelens 2015).

Consider a concrete case. Standing in my backyard, I can see an old bicycle leaning against the wall of my garage, and a potted cactus growing next to it. My perceptual state thus represents various objects: the bicycle, the wall, the cactus, and so on. It also represents them as having a variety of features: the wall as green, the bicycle as leaning at a certain angle, the cactus as positioned at a certain distance from the wall, and so on. All these facts are reflections of the content of my perceptual state. But my current perceptual state also represents its objects as having specific shapes, orientations, and locations; and it represents every other object in my visual field as bearing specific spatial relations to one another and to my vantage point. Considerations like these make it natural to conceive of the contents of visual representations as UNIFIED visual spaces, in the sense that all elements in the space bare relevant spatial relations to one another and to the viewpoint. They form a connected spatial web.

As many authors have noted, visual spaces are distinguished by the fact that they are *perspectival* (see e.g. Budd 1996; Hopkins 1998; Gregory 2013; Casati and Giardino 2013). That is to say that the objects and properties which inhabit visual space are all located there only in virtue of their locational relation to a central perspective or point of view. The viewpoint-relativity of visual space is exhibited in a variety of ways. For example, it arises in relations of depth: the wall of the garage is *further from* my viewpoint than the cactus. And it is reflected in relations of direction: my perceptual state represents the bicycle as to the *left* of my viewpoint, and the cactus as to the *right*. Such directional relations are not limited to crude cardinal directions, but appear to locate every part of every perceived surface in a quantitatively distinct direction relative to my viewpoint.

The most straightforward interpretation of these facts construes visual contents as METRIC VISUAL SPACES. In a metric visual space, every object and property lies at a determinate *distance and direction* from the viewpoint, and thus from each other. 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 such as size, shape, and depth. Metric visual spaces are the sorts of things one could build a physical model of, 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 reconstructing a three-dimensional model of the external world from the retinal input.

A detailed version of the metric visual space idea is developed by Peacocke (1992).[1] Peacocke identifies perceptual contents with SCENARIOS, glossed as "ways of filling out space." A scenario is defined as a coordinate space with a distinguished origin, relative to which properties and objects are assigned definite positions (in specific directions and at specific distances). The origin and coordinate system play roughly the role I've ascribed to viewpoint above.[2]

Yet, for all the idea's simplicity and appeal, visual contents cannot be metric visual spaces. The problem is that visual contents are often indeterminate about essential issues of metric structure. The most prominent example is the perception of depth. It is well-known that perception is often indeterminate with respect to depth: for example, it may record that one object is *behind* another, but not by how much; or it may represent that an object is in some *range* of distances from the viewpoint, but not any particular one. Such indeterminacy is phenomenologically vivid for cases of long distance vision. Consider the perceptual experience one would have looking at the scene illustrated below. We can see that the pyramid in the distance is further from us than the Sphinx, but we have no sense of exactly how far.

[1] See Matthen (2005, pp. 271-289; 2014, pp. 266-279) for another development of the metrical space theory.

[2] Officially, Peacocke's notion of SCENARIO CONTENT is a *set* of scenarios, not a singleton scenario. But the introduction of sets is intended only to model variations in *perceptual acuity*, the degree of clarity or resolution in a perceptual state. (Peacocke 1992, p. 63). Since such variation will not play a major role here, Peacocke's proposal amounts roughly to the claim that visual contents are metric visual spaces of a certain type.

Figure 1: A view of the Great Sphinx of Giza.

Perceptual scientists have documented a wide range of depth cues exploited by the visual system (Solso 1996, ch. 7; Palmer 1999, ch. 5). Some, like objects of familiar size, can be used to estimate absolute relations of depth; in the scene above, this would apply to the perception of the distance from the viewpoint of the rider in the foreground. But others, like occlusion, provide only comparative depth information, as with the perception of the depth of the pyramid in the background. In general, indeterminate depth perception arises when there are sufficient visual clues to determine that one object is further away than another, but insufficient clues to determine how much further.

In addition to these considerations, low-level perceptual representations, computed prior to or independent of depth computations are, perforce, silent on the representation of depth. Potential examples are modular and parallel representations of color, motion, and boundedness (Treisman and Gelade 1980; Treisman 1986; Treisman 1988). Yet these too seem to be clear cases of visual representation.

The metric conception of visual space fails because of its in-built commitments to attributions of determinate depth. Still, the theory seems to correctly characterize the *directional* structure of visual content; for intuitively, it is true that every object in the visual field of a perceptual state is located in a specific viewpoint-centered direction, as metric visual space would predict. The problem with metric visual space is that it treats direction and depth on a par as structural ingredients of visual content. The positive view I'll defend holds fixed the structural role of direction in visual space, but treats the representation of depth, like the attribution of color or texture, as a common but inessential feature of visual representation.

3 Unstructured visual content

If visual contents cannot be thought of as metric spaces, perhaps they should be understood more abstractly as the *set of circumstances at which a visual representation is accurate*. According to the UNSTRUCTURED CONTENT VIEW, visual contents are simply identified with such sets. This is the analogue of the unstructured view of linguistic propositions which identifies them with sets of possible worlds.

Theories of this kind are already familiar in the perception literature, and have some traction in the philosophy of depiction (Ross 1997; Chalmers et al. 2006; Blumson 2009; Brogaard 2011; Abusch 2015). The idea is to identify the content of perceptual states with sets of viewpoint-centered worlds— or sets of pairs of worlds and viewpoints. All and only those pairs of worlds and viewpoints which are compatible with the perceptual state are included in its content. Thus, a percept of a cube from a certain vantage point may include one world containing only the cube, and another world containing the cube as well as a fully occluded sphere, for both are are compatible with the content of the original perceptual state. By contrast, the perception of a cube from close-up, and another of the same cube from further away will express contents that contain the same worlds, but come apart with respect to where they locate the viewpoint in these worlds.

Unstructured theories of content easily accommodate the type of depth indeterminacy discussed in the last section. Consider a perceptual state which represents a mountain as behind a tree, but not how far behind. Its content would be a set of centered worlds; in every such world there would be a visible mountain further from the viewpoint than a visible tree; but the magnitude of the distance would vary from world to world among those in the content. Such worlds would converge insofar as the content is determinate, and diverge insofar as it is indeterminate. Thus the unstructured content view provides a model of visual content that is flexible enough to capture metric indeterminacy, including that of depth.

The central problem facing any unstructured view of visual content is also a familiar challenge for unstructured views of propositional content; this is the problem of *contradictory* contents. On one hand, certain representations seem manifestly to represent content which is contradictory or impossible. On the other hand, since the unstructured view builds contents from possible worlds, as a matter of course it is at a loss to capture the representation of impossible scenarios.

Consider the so-called "waterfall illusion": after visually adapting to a moving surface, such as a waterfall, stationary surfaces, like the ground, appear to be both stationary and moving, thus generating a perceptual contradiction (Crane 1988; Siegel 2016). Or imagine overlapping reflections in a window, which Matthen (2005) suggests cause the percept of incompatible surfaces in the same location. Or consider the case of the perception of perspective drawings themselves. Drawings present conflicting depth cues to the visual system— on one hand, they create a sense of three-dimensional space; on the other, they are clearly recognizable as flat surfaces. The resulting perceptual experience, which Wollheim (1987) called "two-fold," is plausibly yet another case of perceptual contradiction. (Gregory 1970, p. 22)

Contradictory content also arises in depiction. Consider the famous Penrose triangle at left, whose various surfaces manifestly cannot be arranged in the way that our perceptual response to the image seems to require. The same phenomena arises in one of the puzzle drawings of M.C. Escher, below, where a visual contradiction is depicted in the context of an otherwise coherent visual space. In fact, contradictory perceptual content may result from simply *looking* at drawings like these, though empirical evidence remains complex. (Peacocke 1992, p. 74; Schacter et al. 1991)

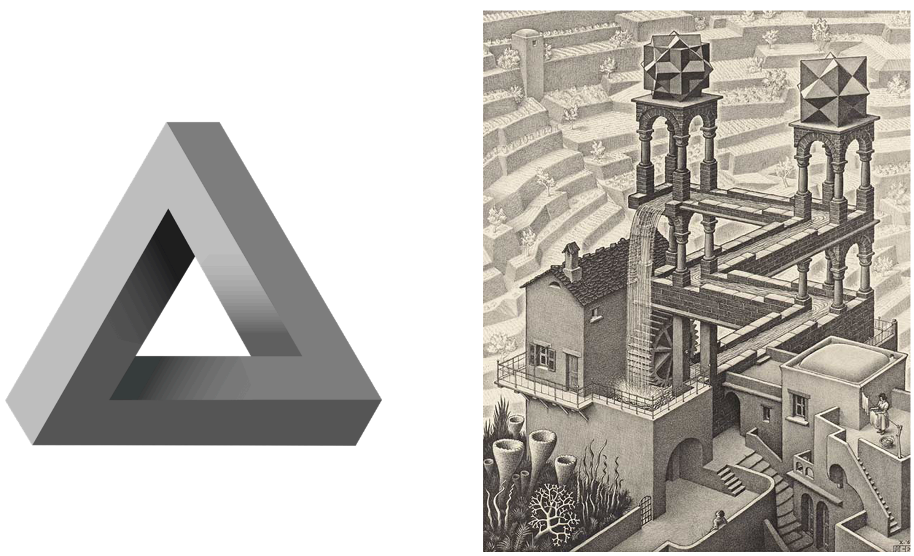

Figure 2: Impossible images: at left, the Penrose triangle (Penrose and Penrose 1958); at right, Escher's *Waterfall* (1961).

All such phenomena appear to pose a basic challenge to the unstructured approach.[3] The problem isn't exactly that the unstructured view assigns contradictory visual representations *no*

[3] Blumson (2009, §VII) has attempted to resuscitate the centered-worlds approach to visual content in the face of contradictory images. But his solution is to think of impossible pictures as expressing sequences of objects and properties, analogues to simple structured propositions. Unfortunately, this move loses the idea that visual representations express spaces of any kind.

content, but rather that it assigns the *same* null content to all contradictory representations, no matter how different the spaces they express. The challenge posed by contradictory visual representations to the unstructured view is that it fails to make distinctions among contents where they ought to be made.

Perhaps some headway can be made on the part of the unstructured view by employing sets of possible and *impossible* worlds, but the problems endemic to such strategies are well known in the philosophy of language (Lewis 1986; Soames 1987). A different tack would be to divide the content of the problematic representations into sets of consistent fragments. But this is a position of last resort, for it would mean giving up on the project of associating contradictory visual representations with the same type of contents as any other visual representation. One may also raise empirical worries about each of the specific cases discussed here— are they really cases of contradictory perception? But at the very least, they confirm the real *possibility* of contradictory visual representations. Visual contradiction should not be ruled out as a matter of conceptual necessity.

Of course, contradictory visual representations are also counterexamples to the proposal that visual contents are metric spaces, a point recognized by Peacocke (1992, p. 74). For no metric space can itself be spatially contradictory. Peacocke's own solution is to introduce an additional layer of content, beyond scenario content, in which subregions of scenarios are associated with "protopropositions" which may be individually or collectively contradictory. As will become clear, my own approach to the problem is aligned with Peacocke's account, modulo the commitment to metric visual space.[4]

The general moral here is that the constituents of visual content cannot themselves be possible metric spaces or worlds, as the unstructured view would have it. Instead, visual contents must be made up of *parts* of such spaces that can in turn be combined into both consistent and inconsistent wholes. In the positive account I'll defend, visual contents are made up of clusters of spatial properties and relations which can be combined consistently or inconsistently in just this way.

4 Perspectival feature maps

4.1 The proposal

In this section I outline an alternative, structured view of visual content, broadly comparable to theories of structured propositions, yet distinctively visual in the structures it assumes. My claim

[4]Peacocke (1992, pp. 75-76) also calls attention to phenomena, such as changes in orientation, where differences in perceptual representation do not make a difference to accuracy conditions. Like visual contradictions, these cases suggest a more granular approach to visual content than is allowed by unstructured theories. Peacocke uses protopropositions to address the issue; the structured theory of visual content I'll defend would work as well.

is that all visual contents take the form of PERSPECTIVAL FEATURE MAPS. Perspectival feature maps are themselves a type of abstract structure which, like Russellian structured propositions, may have concrete objects, as well as properties and relations, as constituents. Which objects and properties get expressed in a perspectival feature map will vary from representation to representation, while the underlying structure will tend to be fixed from cases to case within a given visual modality or system of depiction.

The core structure of a perspectival feature map is a two-dimensional surface, or MAP FIELD, where every point on that surface is associated with a direction that is oriented into the three-dimensional space surrounding it. What makes these structures *perspectival* is the fact that, if we were to locate a viewpoint in the right relationship to the map, every direction in it would converge backwards on this viewpoint. In this way, perspectival feature maps hold fixed the directional organization of visual space, building it into the structure of visual content itself, while eschewing any structural commitment to depth, like that of the metric space theory.[5]

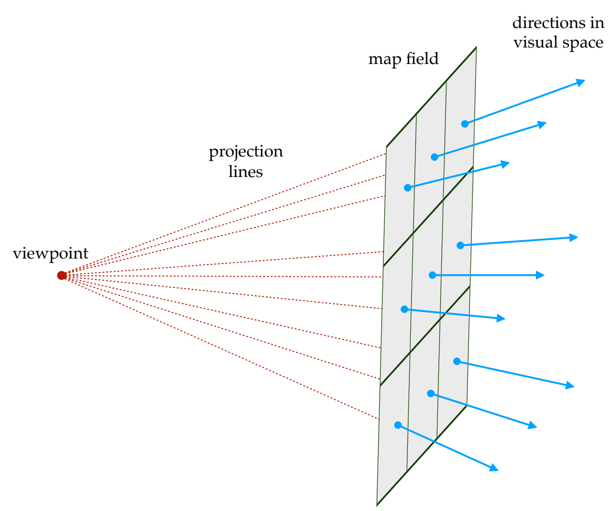

Figure 3: The core structure of a perspectival feature map: the map field and its directions.

Each perspectival feature map defines a kind of directional space— a space whose "dimensions," speaking loosely, are directions emanating from a viewpoint, and whose extent is defined by the shape of the underlying map field. In general, representations that belong to the same visual modality in the same subject will express the same kind of directional space. Differences

[5]A similar picture is anticipated by Matthen (2005, p. 275): "visual directions constitute an omnipresent grid that overlays every scene, indexing the features represented in it. This is an updated 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 does not go on to elaborate the idea of the "omnipresent grid."

in modality— vision vs mental imagery, for example— as well as differences among species, or among systems of depiction, result in different kinds of space. Even within human vision, the character of the underlying feature map will vary between stages of visual processing, and between vision and mental imagery.

Variation in the shape of the map field, as well as the shape, cardinality, and distribution of cells will affect the extent and resolution of the resulting directional space. In human perception, for example, the map field would form a kind of compressed oval, and cells would likely be more densely packed in the center than the periphery, corresponding loosely to the greater focal acuity in the central regions of the visual field. In pictorial representation, the map field is normally flat, rectangular, and uniformly partitioned. (For simplicity, I'll illustrate discussion with flat rectangular map fields, uniformly filled with finitely many non-overlapping cells.)

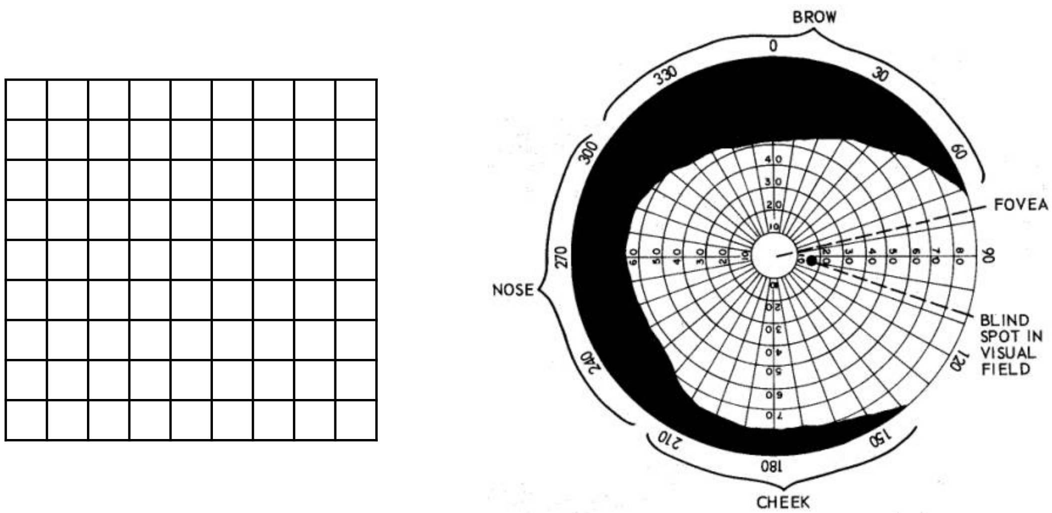

Figure 4: At left, a uniform map field. At right, the graduated map field of monocular human vision (Ruch and Fulton 1960).

Variation in the organization of perspectival directions has a fundamental effect on the internal "shape" of the directional space expressed by a perspectival feature map. Particularly vivid is the contrast between linear perspective and parallel projection systems of depiction. In linear perspective, the direction vectors converge backwards on a point-shaped viewpoint. In parallel projection, the direction vectors are all perpendicular to a projecting plane, itself a notional "viewpoint" which gives rise to the distinctive "god's eye view" associated with parallel projection.

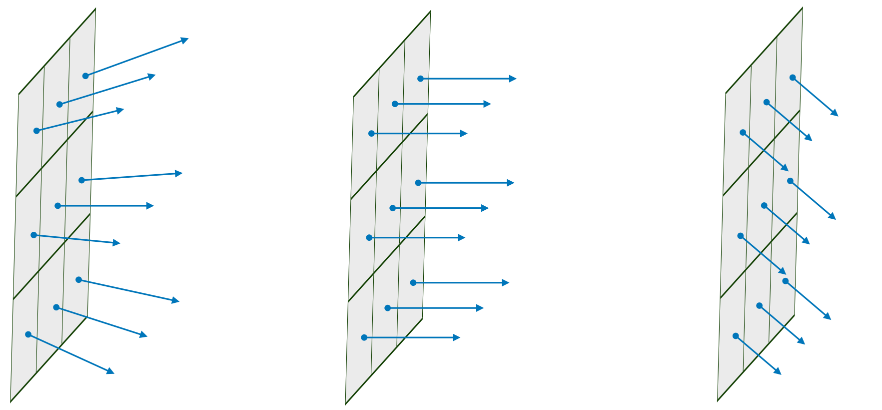

Figure 5: Perspectival map fields based on linear perspective and parallel projection systems of depiction.

To complete a perspectival feature map, regions of the map field are associated with clusters of objects, properties and relations, which I will call FEATURE CLUSTERS. Each feature cluster contains an object and set of properties. For low-level visual representations, typical properties might include surface colors, illumination, motion, and attributes like *edge* and *non-edge*. Higher-level representations may attribute properties of depth, shape, objecthood, and even basic kinds. The objects in feature clusters correspond to a visual representation's singular content, while the properties in feature clusters correspond to its attributive content. Feature clusters are associated with contiguous regions in the map field by a primitive structural relation of LINKING.[6]

[6]For simplicity of exposition, I treat visual contents as including the particular individuals they refer to, in the manner of Russellian propositions. A more sophisticated, Fregean view may be called for, in which the singular elements of visual content are more like senses, modes of presentation, or discourse referents (Burge 2010; Schellenberg 2018).

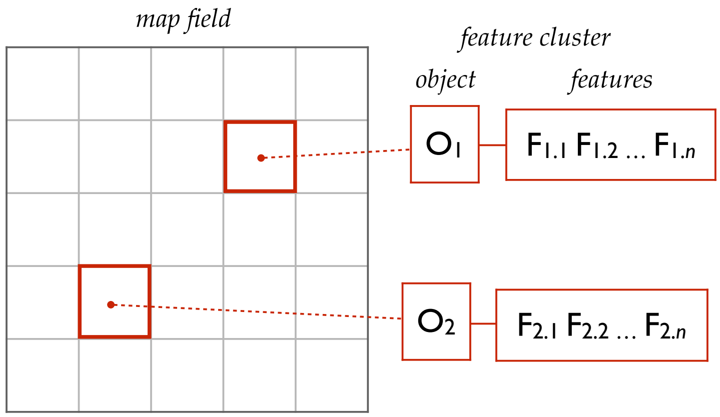

Figure 6: Basic feature map with feature clusters (directions omitted).

In addition to the properties normally associated with visual content, relations also plays a critical role. Candidates include relations between objects like *being the same size as* or *being a darker color than*, and in the next section we will examine relations of depth. But relations cannot be structurally located *within* feature clusters, because feature clusters are associated with single objects, while relations hold between objects. Instead, relations link feature clusters together. (The whole structure now looks a bit like streamers hanging from ceiling tiles.) In this context, the concept of linking is extended to describe a primitive structural relation that connects a sequence of feature clusters to a relation.

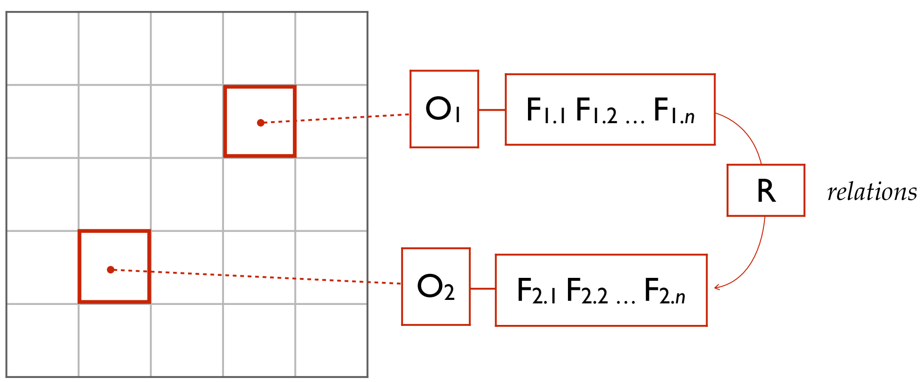

Figure 7: Basic feature map with relations.

Taken as a whole, a perspectival feature map locates each of the objects in its feature clusters in a given direction, and attributes to each its associated properties and relations. A perspectival feature map is accurate when these attributions are correct. More precisely: a perspectival feature map is accurate at a world w and viewpoint v (or centered-world), just in case, when the directions of the map are anchored at v, then for every object in a feature cluster of the map: (i) that object is located in its associated direction in w; (ii) that object instantiates its associated properties and relations in w. In effect, a perspectival feature map displays its accuracy conditions in two-dimensions; it is like a drawing made from bits of reality— constructed not from lines and colors, but actual objects and their features.

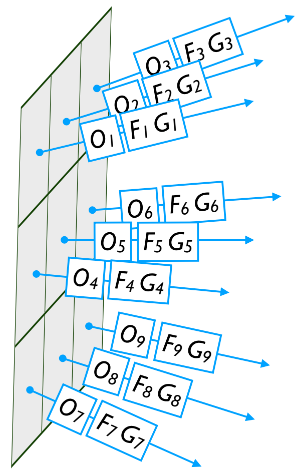

Figure 8: A perspectival feature map with feature clusters.

We are finally in a position to compare perspectival feature maps with structured propositions. Both are object-involving structures that define precise accuracy conditions, or sets of centered-worlds; in this sense both are nominally "propositional." But they differ radically in how they reach this end. At the coarsest level, structured propositions are basically tree-like structures, while perspectival feature maps are basically array-like. The primitive structural relation in a structured proposition (depending on specifics) is something like a triadic relation between ordered sibling nodes and their parents, whereas the building blocks of a perspectival feature map are the geometrical relations that define the map field and its directions. And they integrate concrete constituents in different ways: in structured proposition, each leaf is associated with a single object, property, or relation, whereas perspectival feature maps are populated with clusters of elements, not individuated leaves. In this more specific sense, perspectival feature maps are not "propositional" at all, but exhibit an equally legitimate alternative structural form. In the end, propositional structure is revealed to be just one kind of structure which can inhabit the semantic role of content. (Byrne 2001, pp. 201-2; Crane 2009; Grzankowski 2015; Camp 2018).

The claim that visual contents are perspectival feature maps is intended to get at an essential mark of visual representation: it is part of what it is to be a visual representation that it express content with the structure of a perspectival feature map. Not all representations with graphical structure are visual, because not all express the relevant type of content. Venn diagrams, for example, as well as other many other diagrams and graphs, are not themselves visual representations by present lights, because the content of a Venn diagram is a set of logical relationships, not an arrangement of objects and properties organized in a directional array.

Furthermore, representations whose content is merely spatial will not necessarily also be visual. An allocentric description of the environment, for example, might represent the location of various landmarks within a fixed coordinate system (Camp 2007, p. 158). But if it does not represent its subject matter via viewpoint-centered directions on a visual field, it is not visual representation in my sense. This might be so even for spatial representations that directly interact with the visual system. So-called "object-centered" representations define the intrinsic shapes of objects, without relativization to an external viewpoint (Marr 1982). (A very simple example would be the description: *object x is a 2-inch diameter sphere.*) Object-centered representations seem to play a role in perceptual processing, but they are not, by present lights, visual representations themselves, because their content is not perspectival. Within the more general class of spatial representations, only those which provide viewpoint-centered descriptions of the environment belong to the natural class of visual representations.

4.2 Depth indeterminacy and visual contradiction

The virtues of the perspectival feature map analysis are brought out in application to the problem cases of depth indeterminacy and contradictory visual content, faltering points for the alternative accounts of visual content discussed above.

To begin, we should distinguish two kinds of features that may appear in a feature map. Features may include non-relational properties like *sphere* or *cube*; but they may also include what many have termed PERSPECTIVAL PROPERTIES: properties of objects which are only defined relative to a viewpoint. Absolute depth properties are a prime example, effectively attributing to an object the property of being a given distance away *from the viewpoint*. In the present framework, I interpret the relevant aspect of the viewpoint to be the surface of the map field itself. Besides perspectival properties, there are also perspectival relations— relations between represented objects relative to a viewpoint. Attributions of relative depth have this form; thus the relation *more distant than* holds in the first place between two objects, but makes implicit reference to the viewpoint.

Now consider the original case of indeterminate depth perception (and depiction) illustrated by the visual relationship between the Sphinx and the pyramid behind it. Both the Sphinx and the pyramid are associated with distinct feature clusters. Perhaps the Sphinx is even assigned a determinate depth relative to the viewpoint. But there is not enough visual information to assign the pyramid a determinate depth from the viewpoint. Instead, it is related to the Sphinx by the *more distant than* relation, and it is this relation which appears in the corresponding perspectival feature map.[7] Metric indeterminacy is easily accommodated by including only such relational depth features and no absolute ascriptions of depth. The proposed analysis is illustrated in Figure 9; here the diagram of the feature map is overlaid on the visual representation that expresses it. I include the attributive *3D solid* for illustration, and suppress the visualization of directions.

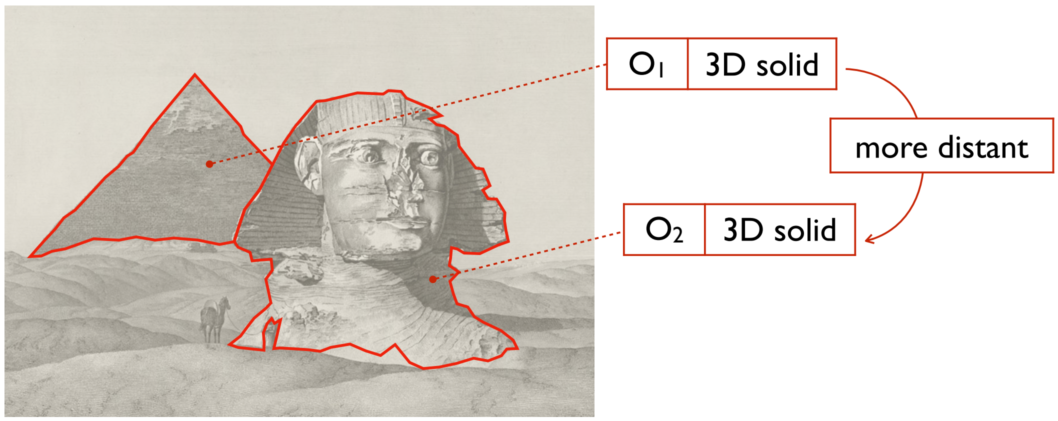

Figure 9: Indeterminate depth relations in a feature map.

The representation of indeterminate depth was a problem for metric visual space theory, which built metric depth relations into the structure of visual content. By contrast, while visual direction is a fixed aspect of the structure of perspectival feature maps, depth is treated like just another optional feature. Consequently, the representation of depth may be absolute in some cases, indeterminate in others, and wholly absent in still others.

For broadly analogous reasons, perspectival feature maps are well suited for capturing contradictory content in a way that possible-worlds accounts are not. This stems from the fact that there is no consistency requirement on feature clusters. In principle, the same feature cluster could contain both the properties F and $\neg F$. Indeed, this sort of content is plausibly associated with the Waterfall Illusion where the same surface may simultaneously be attributed the properties *stationary* and *moving*.

In the case of the Penrose triangle— assuming that it, or the percept it causes has impossible content— the contradiction is less direct. A careful analysis of the contradiction here is probably best formulated in terms of a system of line labeling like that developed by Huffman (1971) and others. But very roughly, three incompatible spatial relations are simultaneously attributed. We may say that two flat surfaces stand in the relation of *convexity* when they form a convex

[7] Cian Dorr (personal communication) raises the following concern: perspectival feature maps require that relations of relative distance are asymmetric; but these can be expressed either by the *further than* relation or by the *closer than* relation. Yet intuitively, visual content does not distinguish such attributions. It would seem a vice of the present model that it imposes a choice here— a problem avoided by the metric space approach, where depth relations are encoded directly in metric structure.

edge, relative to the viewpoint. Labeling the three facing surfaces in the figure O_1, O_2, and O_3, then the image simultaneously expresses three relations: $convex(O_1, O_2)$, $convex(O_2, O_3)$, and $convex(O_3, O_1)$. But, as Huffman shows, by the content of $convex$, and the laws of spatial geometry, these three relations can never be simultaneously satisfied by three flat surfaces. Hence the picture is contentful, yet its content is contradictory.

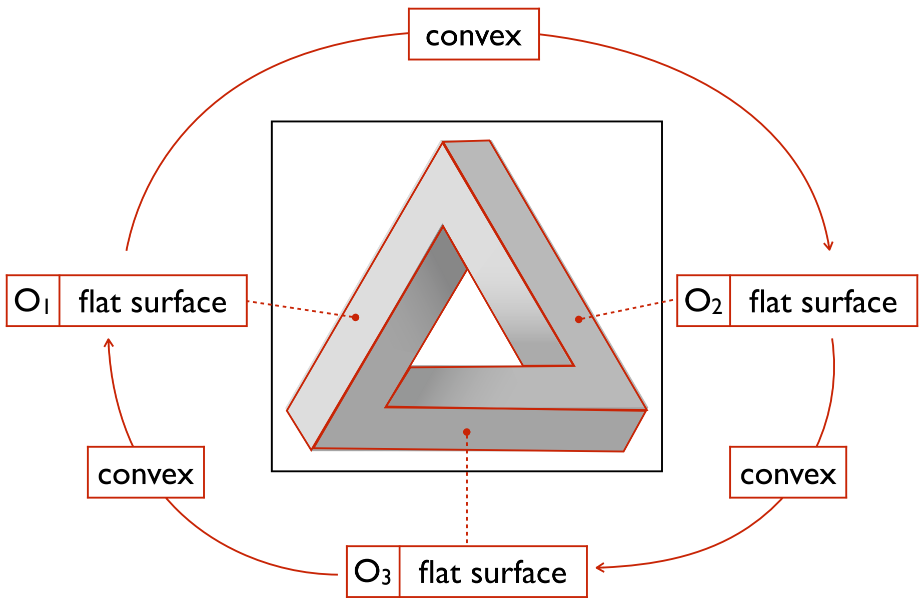

Figure 10: Inconsistent spatial relations in a feature map.

In short, by breaking content down into structured feature clusters, the present proposal allows contradictions to arise both among and within feature clusters. Yet by arranging feature clusters into a unified directional array, we maintain the sense in which visual contents comprise visual spaces.

4.3 Visual field and visual direction

The core structure of a perspectival feature map is made up of two elements: a two-dimensional map field, the abstract counterpart of the visual field; and a suite of directions associated with points on the map field. These ingredients in turn reflect two immutable facts about visual content. First, every object and property involved in visual content is represented via a region of the visual field. And second, every such object and property is located along a viewpoint-centered direction.

The primary role of the plane of the map field is to constrain the scope of visual content. Visual representations do not normally locate objects in all directions, but only within a window of a predefined shape. And they do not normally locate object at depths, but only those which lie *in front* of the plane of this window. The map field encodes this window as a structural limit on content. Objects that lie outside of the directional window of the map field, or behind it, are not represented. In addition, through division into cells, map fields providing visual content with a base-line level of acuity or resolution.

The second aspect of the core structure of perspectival feature maps is the distribution of directions across the field. A consequence of this commitment is that visual representations are never indeterminate with respect to direction, though they may be indeterminate with respect to any other property. To be sure, a map field with large cells will locate its feature clusters within comparatively large directional ranges, thus incurring a kind of indeterminacy. Still, the level of determinacy, and any variation in determinacy is always fixed by the structure of the map field. Visual contents are always determinate with respect to direction up to the level of acuity allowed by the map field in question. Direction is not like depth in this respect, which may vary across the visual field from one part of a representation to another.[8]

The core structure of perspectival feature maps corresponds to those aspects of visual content which, roughly, are fixed across visual representations. What is constant across visual episodes is the directions in which objects are located and the shape of the visual field through which they are located. What is variable are the particular objects perceived and the non-directional features, including depth, associated with them (Matthen 2005, pp. 274-276). Perspectival feature maps encode what is essential to visual content in their core structure, relegating what is contingent to the feature clusters.

The fact that visual field and visual direction are essential and stable ingredients of visual content reflects, in broad brushstrokes, the function of perceptual computation: perception, it is thought, has the function of reconstructing a representation of the external world on the basis of incoming light. Since light arrives at the retina in angularly arranged straight lines, computations which attempt to estimate the location of the distal source of retinal illumination will inevitably have to trace that source back along the directional lines on which its reflection arrived. Failure to do so would miss the most basic source of environmental information available to the eye. Thus, we should expect directional information to be the structural basis for any further elaboration of visual content.

The distinction between core and variable structure also reflects different ways that content is encoded in the course of visual computation.[9] Features like color, shape, or motion are thought to be explicitly represented, the output of so-called "feature detectors." Meanwhile, the directional location associated with any such feature detector is typically assumed to be implicit, a product of the functional architecture that connects the feature detector to the retina. Content implicitly

[8] Although the perceptual system seems to make occasional errors about attribution of direction, there are no clear examples in which it is not represented at all (Treisman and Schmidt 1982; Ashby et al. 1996; Pylyshyn 2006, pp. 177-8).

[9] Thanks to William Kowalsky for help with this point. See also Matthen (2005, p. 276).

encoded in the fixed causal structure of the perceptual system is captured in the perspectival feature map's core structure, while explicitly represented content is reflected in its variable structure. A similar line of reasoning seems to apply in the linguistic case. There, the contents explicitly encoded by individual words show up as concepts, functions, or individuals in the nodes of structured propositions. Meanwhile, the implicitly encoded composition of these elements is reflected in the structure of the propositions themselves.

4.4 Feature maps in perceptual science

The theory of content defended here is inspired by the use of feature maps in perceptual science, which date back at least to the work of Marr (1982) and Treisman (1988), and are now ubiquitous in vision science and computer vision.[10] Feature maps are invoked both as as data structures in algorithmic descriptions of visual processing, and as physical structures at the level of neural implementation. (Frisby and Stone 2010, ch. 10) They are nearly always deployed to understand the representations involved in a specific visual subsystem, like edge or color detection, rather than as a means of unifying the content of such subsystems, which is my aim here.

In addition to differences of scope, the present proposal should be distinguished from the feature maps that figure in the scientific literature in several ways. (i) Feature maps are typically construed as descriptions of representational *vehicles* (populated with *symbols*), whereas perspectival feature maps are a type of *content* (populated with *objects* and *properties*). (ii) Traditional feature maps focus exclusively on the *features* attributed (like orientation or color), where as perspectival feature maps are specifically designed to combine features with *objects* to determine accuracy conditions. (iii) What is most distinctive about the present approach is the explicit inclusion of the directional array as part of the core structure of the feature map. Almost all uses of feature maps either make no mention of visual direction or treat it as implicitly specified by a "line of sight" (Marr 1982, p. 283; Tye 2000, p. 81; but see Matthen 2005, pp. 274-6), with no indication of the role of direction in content. Perspectival feature maps recast direction as a central component of content and a determiner of accuracy conditions.

5 Equivalent structures

Perspectival feature maps are not the only kind of structure that can carry visual information. In this section, I consider two informationally equivalent structures, one based on purely directional structure, the other propositional, and explain why perspectival feature maps are still the preferred account of visual content.

According to a PURELY DIRECTIONAL approach, visual contents are identified with a bundle

[10] The terminology of "feature maps" comes from the work of Treisman (1980; 1986; 1988). She uses the term to refer to a specific kind of representation of a single determinate feature across the visual field. Treisman's feature maps (qua representations) do express feature maps in my sense (qua content), but they are not the only thing that does.

of directions, emanating from an origin point, each of which is associated with a feature cluster. No additional role is assigned to an intervening, two-dimensional surface. Such a structure can be thought of as derivative of Peacocke's notion of a scenario, by removing the imposition of depth relations and preserving direction, though this formal amendment would require a deeper revision to Peacocke's conception of perceptual content as a way of "filling out space."

By doing away with the map field, however, the purely directional approach looses the substantive constraints which map fields impose on visual content. Recall that map fields introduce limitations on the size, shape, and location of the bundle of directions expressed in visual content, as well as the underlying cell structure. These constraints capture what is common across contents of a given modality. In visual perception, for example, the contents expressed always locate objects within a visual field of the same roughly-oval shape. Perspectival feature maps simply reify these constraints in the structure of content. The same generalizations may be regained by the pure direction theorist through explicitly stipulated constraints on the available directions. Still, from a methodological perspective, the way of perspectival feature maps is preferable, for it treats the essential features of visual content uniformly, building them all into structure, rather than putting some into structure and some into local axioms, as the purely directional approach would have it.

In addition to the purely directional approach, there are also equivalent structures which are straightforwardly propositional. Though feature maps are not propositional structures, it should be recognized that propositions can, in some important sense, express the same information as perspectival feature maps. This is to be expected, given the general availability of algebraic representations of geometric structures. By representing feature clusters and directions with simple propositions, and then conjoining them, the feature map as a whole may be translated into propositional form. When a proposition can be translated from a feature map in this manner I will say that the proposition SPECIFIES the feature map. Given the availability of such propositional translations, why attribute feature map structure to visual contents in the first place, and not the more familiar propositional structure? Several reasons prevail.

First, the tree-structure of traditional propositions has no special philosophical status; it makes sense as an account of the content of sentences, precisely because sentences themselves have such tree-structure, and different trees seem to mark differences in linguistic meaning. But for representations which are not in any straightforward sense language-like— think especially of pictures, and low-level retinotopic perceptual representations— there seems to be no clear reason to expect that their content would exhibit a tree-like structure.

Second, not all propositions specify perspectival feature maps. Even if visual contents were propositional, they would correspond to only those highly specialized propositions which happen to specify perspectival feature maps. Given that perspectival feature maps seem to be playing the more basic explanatory role here, it seems we should recognize them as directly characterizing the structure of visual content. Visual content may be described using propositional structures, but these are ultimately redescriptions of the underlying visual form.

Finally, a given perspectival feature map can be specified by many different propositional structures. This can be achieved, for example, simply by changing the order of conjuncts in the relevant proposition (or introducing other variations in formulation that are logically equivalent). But these differences in structure do not correlate with differences in visual content. As a result, as an account of visual content, propositions appear to be too fine-grained; they mark structural distinctions where there are none. By contrast, perspectival feature maps are able to more nearly reflect all and only the differences actually found in visual content.

That said, there are special cases where a representation clearly seems to consist largely of complex propositional structures, but is nevertheless visual. Consider the output of sophisticated computer vision software, like that used in the visual modules of self-driving cars.[11] The data structures produced are written in an algebraic code, but still describe the directions of objects and properties relative to a viewpoint. Despite their obvious differences with pictures and retinotopic representations, such codes express content that is distinctively visual. Furthermore, we do not at the present time know very much about the implementation of late vision; it's representations might well resemble computer code more than retinotopic maps.

Consider a long sentence in a predicate calculus which specifies a perspectival feature map. What is the content of such a sentence? Given its linguistic format, the standard reasons apply for considering its content to be a structured proposition. What then should we say about its visual content? Here, I think, we may postulate multiple levels of content, just as multiple levels of representation are commonly recognized in computational theory. (Marr 1982; Newell 1982; Pylyshyn 1986) We may speak both of the propositional content and the visual content of a given representation. Determining which way of speaking is appropriate depends on our explanatory purposes. When thinking of the sentence in terms of its function to describe the directional layout of the environment– a function shared, more or less, with perception and pictorial representation– visual content is the relevant object of inquiry. But when thinking of what it entails for the purposes of logical proof, for example, propositional content may be more relevant. In ascending to a level of abstraction in which we think of such representations as visual representations, as opposed to segments of code, we abstract away from possible differences in the underlying substrate, and focus on the representation's distinctively visual function.

[11] 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, and the picture may express visual content, the underlying code merely represents the picture itself, not the picture's content.

6 Stages of visual processing

I conclude by applying the perspectival feature map analysis to recent empirical hypotheses about visual perception and mental imagery. A prominent line of research in perceptual psychology, stemming from the work of Treisman (1980; 1986; 1988), Pylyshyn (2003; 2007), and others, conjectures two rough stages in perceptual processing. In the first stage, low-level visual features are detected and registered uniformly across the visual field. Distinct representations are posited for the detection of different feature dimensions; candidate dimensions include shape, color, orientation, boundendess, and motion, among others. Treisman called these *feature maps* because they register the presence of features in a map-like array covering the visual field. At the second stage, these features are integrated into small set of *object files*. Object files track the location of mid-sized objects, and collect together in a single accessible frame all of the features associated with those object by earlier stages of visual processing. Such object files are thought to be the result of comparing and collating the low-level Treisman-feature maps, and identifying which clusters of detected features in the visual field likely correspond to larger objects. The object file system is thought to be highly restricted, representing only 3 or 4 objects at a time (Green and Quilty-Dunn 2017).

Both types of representation are well-suited to analysis in terms of perspectival feature maps, though the kind of perspectival feature map expressed in each case is quite different. To begin, how should we understand the content of a typical Treisman-map? My account requires that every feature cluster contain an object.[12] While singular representation does not play a central role in the scientific theory here, only relatively small entities, like edges, patches, or parts of surfaces would typically instantiate the corresponding low-level features, so these will form the basis of the feature clusters. The perspectival feature map for a Treisman-map will then have a number of distinctive characteristics. It will contain a feature cluster for every cell or small group of cells. And for a given map, the feature clusters will contain a single type of object— all edges or surface patches, for example. Further, the feature clusters will be uniform, including the same type of property across the map field. And each feature cluster will typically contain only one feature, the relevant particular feature from the appropriate feature dimension.

Content at the post-integration object file stage looks quite different. Since only a small number of object files are ever maintained at once, the corresponding feature map may contain only a small number of total feature clusters Each feature cluster will be associated with a relatively large region of the map field— corresponding to the retinal image of the perceived object. Those feature clusters which are included will have something like familiar mid-sized volumetric objects, or Spelke-objects, as singular content (Spelke 1990; Green 2018). And these feature clusters

[12] By contrast, Clark (2006), for example, holds that, at low-levels, features are attributed to locations. See Matthen (2004) and Cohen (2004) for reasons to favor an object-based approach.

will contain a number of different features, the result of integrating features at a lower level. Over time, the same feature cluster may be associated with different segments of the map field. On this scheme, feature clusters themselves are the contents of object files; object files contribute to overall visual content via their association with regions of the map field.

Although each type of representation is compatible with the perspectival feature map analysis, the two examples suggest very different kinds of visual content. Treiseman-feature maps express perspectival feature maps made of a huge number of uniformly distributed feature clusters each of which contains a single property; representations at the object-file stage express perspectival feature maps with only a few feature clusters, each of which will contain many properties.

The latter representations are "pictorial" in their gross spatial organization, but not "pictorial" in the sense that they do not uniformly describe all details of the visual space. This feature of seems to be a characteristic of many high-level processes in the visual system. For example, mental imagery is widely thought to be "patchy"– fully picture-like in some details, but gappy or descriptive in others. In an example made famous by Dennett (1969), when instructed to visualize a tiger, many will experience mental imagery that is vividly detailed in certain respects— perhaps the motion of the tiger or its outline are clear— but few will precisely represent the number of visible stripes. In the minds eye, as it were, the tiger is simply "striped." Partly in response to this problem, Tye (2000) proposes to analyze mental images as "symbol-filled arrays," feature maps at the level of representational vehicle. Tye's thought is that, in some areas of a mental image, the shape of the array, together with detailed symbolic enrichment, may give rise to vivid, picture-like images. But in other areas, only descriptive or summary symbols are assigned to broad regions of the array. This variation gives rise to the patchy quality of mental imagery.

Tye's basic strategy can be smoothly translated to the framework of perspectival feature maps. Relatively pictorial aspects of a mental image express regions of the feature map that are densely packed with feature clusters, rich in spatial properties and relations. Relatively patchy or descriptive aspects of an image express feature map regions that are are sparse with feature clusters. In the case of Dennett's tiger, a large segment of the feature map might be associated a single surface in the object position, and the property of *being striped* in the feature position. Meanwhile, other segments may include individual stripes, their colors, and locations.

Similar phenomena arise in certain kinds of gappy or selective drawing systems, such as sketches. And comparable results can certainly be imagined for the data structures of computer vision. From a technical perspective, what these cases illustrate is that not all cells of a map field need be recruited as bearers of feature clusters. When a region carries a feature cluster, but its sub-regions do not, the resulting content will be relatively patchy and selective. The results are still visual contents, albeit ones that begin to depart from the pictorial norm.

References

Abusch, Dorit (2015). "Possible worlds semantics for pictures". In: *Blackwell Companion to Semantics*. Ed. by Lisa Mathewson et al. Forthcoming. Wiley, New York.

Ashby, F Gregory et al. (1996). "A formal theory of feature binding in object perception." In: *Psychological review* 103.1, p. 165.

Blumson, Ben (2009). "Pictures, Perspective and Possibility". In: *Philosophical Studies* 149.2, pp. 135–151.

Brogaard, Berit (2011). "Centered worlds and the content of perception". In: *A Companion to Relativism*, pp. 137–158.

Budd, Malcolm (1996). "How pictures look". In: *Values of art: Pictures, poetry, and music*. Blackwell.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

Burge, Tyler (2005). "Disjunctivism and perceptual psychology". In: *Philosophical Topics* 33.1, pp. 1–78.

— (2018). "Iconic Representation: Maps, Pictures, and Perception". In: *The Map and the Territory*. Springer, pp. 79–100.

Byrne, Alex (2001). "Intentionalism defended". In: *The Philosophical Review* 110.2, pp. 199–240.

Camp, E. (2007). "Thinking with Maps". In: *Philosophical Perspectives* 21.1, pp. 145–182.

— (2018). "Why Maps Are Not Propositional". In: *Non-Propositional Intentionality*. Ed. by A. Grzankowski and M. Montague. Oxford.

Casati, Roberto and Valeria Giardino (2013). "Public Representation and Indeterminicies of Perspectival Content". In: *Enacting Images*. Ed. by Zsuzsanna Kondor. Herbert von Halem Verlag, pp. 111–126.

Chalmers, David et al. (2006). "Perception and the Fall from Eden". In: *Perceptual experience*, pp. 49–125.

Clark, Austen (2006). "How do feature maps represent?" In: *Early Content Conference, University of Maryland*. Vol. 22.

Cohen, Jonathan (2004). "Objects, places, and perception". In: *Philosophical Psychology* 17.4, pp. 471–495.

Crane, Tim (1988). "The waterfall illusion". In: *Analysis* 48.3, pp. 142–147.

— (2009). "Is perception a propositional attitude?" In: *The Philosophical Quarterly* 59.236, pp. 452–469.

Dennett, D (1969). *Content and consciousness*. Routledge & Kegan Paul.

Erkelens, Casper J (2015). "The perspective structure of visual space". In: *i-Perception* 6.5, p. 2041669515613672.

Frisby, John P and James V Stone (2010). *Seeing: The computational approach to biological vision*. The MIT Press.

Green, Edwin James and Jake Quilty-Dunn (2017). "what is an object file?" In: *The British Journal for the Philosophy of Science*.

Green, EJ (2018). "What Do Object Files Pick Out?" In: *Philosophy of Science* 85.2, pp. 177–200.

Gregory, Dominic (2013). *Showing, sensing, and seeming: Distinctively sensory representations and their contents*. Oxford University Press.

Gregory, Richard Langton (1970). *The Intelligent Eye*. McGraw-Hill.

Grzankowski, Alex (2015). "Pictures have propositional content". In: *Review of Philosophy and Psychology* 6.1, pp. 151–163.

Haugeland, J. (1991). "Representational Genera". In: *Philosophy and Connectionist Theory*. Ed. by W.M. Ramsey, S.P. Stich, and D.E. Rumelhart. Erlbaum.

Hopkins, Robert (1998). *Picture, Image and Experience: A Philosophical Inquiry*. Cambridge University Press.
Huffman, D.A. (1971). “Impossible Objects as Nonsense Sentences”. In: *Machine intelligence*, p. 295.
King, Jeffrey C. (2017). “Structured Propositions”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2017. Metaphysics Research Lab, Stanford University.
Koenderink, Jan and Andrea van Doorn (2008). “The structure of visual spaces”. In: *Journal of Mathematical Imaging and Vision* 31.2-3, p. 171.
Lande, Kevin J (2018). “The Perspectival Character of Perception”. In: *The Journal of Philosophy* 115.4, pp. 187–214.
Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.
Luneburg, R. (1947). *Mathematical Analysis of Binocular Vision*. Princeton University Press.
Marr, David (1982). *Vision*. New York, NY: Henry Holt and Co.
Matthen, Mohan (2004). “Features, places, and things: Reflections on Austen Clark’s theory of sentience”. In: *Philosophical Psychology* 17.4, pp. 497–518.
— (2005). *Seeing, doing, and knowing: A philosophical theory of sense perception*. Clarendon Press.
— (2014). “Image Content”. In: *Does Perception Have Content?*, pp. 265–90.
Millar, Boyd (2016). “Frege’s Puzzle for Perception”. In: *Philosophy and Phenomenological Research* 93.2, pp. 368–392.
Newell, A. (1982). “The knowledge level”. In: *Artificial intelligence* 18.1, pp. 87–127.
Palmer, Stephen E (1983). “The psychology of perceptual organization: A transformational approach”. In: *Human and machine vision* 1, pp. 269–339.
Palmer, Stephen E. (1999). *Vision Science: Photons to Phenomenology*. The MIT Press.
Peacocke, Christopher (1992). *A Study of Concepts*. The MIT Press.
Penrose, Lionel S and Roger Penrose (1958). “Impossible objects: A special type of visual illusion”. In: *British Journal of Psychology* 49.1, pp. 31–33.
Pylyshyn, Z. (2006). *Seeing and Visualizing: It’s Not What You Think*. Cambridge, MA: MIT Press.
Pylyshyn, Zenon (2003). *Seeing and Visualizing: It’s not what you think*. Cambridge, MA: MIT Press.
Pylyshyn, Zenon W (2007). *Things and places: How the mind connects with the world*. MIT press.
Pylyshyn, Z.W. (1986). *Computation and cognition: Toward a foundation for cognitive science.* The MIT Press.
Rogers, Sheena (1995). “Perceiving pictorial space”. In: *Perception of space and motion* 5.
Ross, Jeff (1997). *The Semantics of Media*. Kluwer Academic Publishers.
Ruch, TC and JF Fulton (1960). *Medical Physiology and Biophysics*. B. Saunders Co.
Schacter, Daniel L et al. (1991). “Implicit memory for possible and impossible objects: constraints on the construction of structural descriptions.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17.1, p. 3.
Schellenberg, Susanna (2018). *The Unity of Perception: Content, Consciousness, Evidence*. Oxford University Press.
Siegel, S. (2011). *The Contents of Visual Experience*. Oxford University Press.
— (2016). “The Contents of Perception”. In: *Stanford Encyclopedia of Philosophy*.
Soames, Scott (1987). “Direct Reference, Propositional Attitudes, and Semantic Content”. In: *Philosophical Topics* 15, pp. 47–87.
Solso, Robert L (1996). *Cognition and the visual arts*. MIT press.
Spelke, Elizabeth S (1990). “Principles of object perception”. In: *Cognitive science* 14.1, pp. 29–56.
Suppes, Patrick (1977). “Is visual space Euclidean?” In: *Synthese* 35.4, pp. 397–421.

Treisman, Anne (1986). "Features and objects in visual processing". In: *Scientific American* 255.5, 114B–125.
— (1988). "Features and objects: The fourteenth Bartlett memorial lecture". In: *The quarterly journal of experimental psychology* 40.2, pp. 201–237.
Treisman, Anne and Hilary Schmidt (1982). "Illusory conjunctions in the perception of objects". In: *Cognitive psychology* 14.1, pp. 107–141.
Treisman, Anne M and Garry Gelade (1980). "A feature-integration theory of attention". In: *Cognitive psychology* 12.1, pp. 97–136.
Tye, Michael (2000). *The imagery debate*. The MIT Press.
Wagner, Mark (2012). *The geometries of visual space*. Psychology Press.
Wollheim, Richard (1987). *Painting as an Art*. Thames and Hudson.