The Structure of Visual Content

Gabriel Greenberg  |  January 7, 2019  |  Department of Philosophy, UCLA

This essay defends a structured theory of visual content, analogous to structured theories of propositions from the philosophy of language. Visual content takes the form of what I will call a *perspectival feature map*. The proposal builds on work in perceptual science and philosophy that analyzes visual representation in terms of *feature maps* (e.g. Marr 1982, Treisman 1988, Tye 2000, Matthen 2005, Clark 2006, Lande 2018), two-dimensional arrays populated with representations of features in the environment. Perspectival feature maps develop this idea in a novel way, adding a distribution of viewpoint-centered directions into the structure of the array itself, and allowing that regions the array are linked to objects, properties, and relations. Such structures capture the essential character of visual content. The claim is meant to apply across visual modalities, including visual perception, mental imagery, perceptual memory, and pictures.

I motivate the structured account of visual content in part by showing how other, more familiar proposals face basic challenges. In Section 2, I argue that theories which identify visual contents with concrete, spatial scenarios (e.g. Peacocke 1992) are unable to capture important dimensions of spatial indeterminacy. And, as I argue in Section 3, 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, I argue, addresses both concerns. Section 4 lays out the key commitments and motivations of the perspectival feature map approach. Section 5 goes on to develop an account of the structure of perspectival feature maps, while Section 6 provides an explicit definition of accuracy. Sections 7 and 8 examine the consequences of these ideas for the nature of visual content and its relationship to other kinds of structured content. Finally, in Section 9, I discuss extensions and applications of the analysis to a variety of phenomena in visual perception and mental imagery, including feature detection, object files, and patchy imagery. Section 10 is a conclusion. In the appendix I provide a formal definition of perspectival feature maps and their accuracy conditions.

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. It covers not only visual perception,
but also mental imagery, perceptual memory, as well as man-made pictures. The conjecture of this essay is that all members of this diverse group express a common type of content, **visual content**. My aim is to give a general account of the structure of visual content.

The original case of visual content is the content of visual perception, and this will be my primary case study in the discussion to follow. I will 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 proposal developed here is intended to be general enough to include the content expressed at each of these points.\(^1\)

Beyond perception, visual representation includes both mental imagery and visual memory, even though these are invariably more schematic than perception. 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. On the face of it, despite the enormous diversity in the function and physical realization of these representations, all seem to express content of a characteristically visual type: at the very least, all seem to represent various objects and features arranged in space, organized around a central viewpoint or perspective. This does not amount to a definition of visual representation, but it is enough of a description, I hope, to indicate the class of distinctively visual representations I have in mind.

Because the variety of visual representations is wide, any account of their commonalities must, perforce, be thin. I will have something to say about the sense in which all visual representations express contents which are spatial, perspectival, and bound by a visual field, and I’ll offer specific proposals about how the objects and properties which figure in these contents are organized. But I’ll be intentionally non-commital about many of the urgent topics of contemporary philosophy of perception, such as which properties in particular, and what types of singular content, are represented in vision (Siegel 2011). My proposal helps to clarify these debates, but it does not resolve them.

In what follows I develop a structured account of visual content, as opposed to an unstructured account, which would treat visual contents as sets of centered-worlds (or the like). The distinction is analogous to one familiar from philosophy of language: that between **structured propositions**— understood as tree-like (or sentence-like) structures assembled from senses, or objects and properties— and **unstructured propositions**, considered as sets of possible worlds (King 2017). The

\(^1\)In particular, I mean to include both late-stage perceptual representations, where the environment is divided into objects and properties, as well as early-stage “feature maps” which reflect the distribution of single features, or feature dimensions (Treisman 1988).
particular kind of structure I propose for visual content is that of a perspectival feature map, a type of two-dimensional array populated with objects, properties, and relations. In broad generalization then, structured propositions have a tree-like structure, while structured visual contents, as I conceive them, have an array-like, or feature-map structure.

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; the latter claim may also be true, but I don’t take a stand on it here. 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 and structure of representational vehicles. The issue of whether all visual representations share a common type of vehicular format is complicated by the wide range of physical media in which visual representations may be realized: consider the differences in implementation between pictures, computer images, and perceptual states at various levels of retinotopy. It may be possible to argue that all such representations share a common vehicular format, but this is a difficult question which I prescind from here. Instead, my claim is that visual representations, whatever their physical realization or vehicular structure, express a common type of content.

In the presentation to follow, I advocate for one account of visual content over others. But I am often inclined to a more pluralistic attitude: perhaps a given representation may be characterized as expressing different kinds of content at different levels of explanation. I find it plausible, for example, that linguistic representations have both structured and unstructured propositional content, corresponding to different levels of abstraction. Similar considerations may apply to the visual case. If so, perhaps visual representations can be appropriately characterized in terms of metric visual spaces at one level (the subject of Section 2), sets of centered worlds at another level (the subject of Section 3), and perspectival feature maps at still another. This is not exactly what I argue for below, but my conclusions here can be reasonably translated into a more ecumenical framework. Where I argue that visual content must have such and such properties, the reader may take me to be arguing that there is a level of visual content that must have such and such properties. Either way, it is not my intention to argue that the alternative accounts of content are without explanatory significance—only that perspectival feature maps are a necessary component of the explanation of visual content, one that covers a range of cases and dimensions of variation not accounted for by the alternatives.

§1 Visual content

[2]Those who reject a purely structural level of representational vehicles may instead interpret these debates as arguments about the structure of content itself, in which case my thesis is directly apropos.
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).

To illustrate, consider a concrete case. Standing in my backyard, I can see an old bicycle leaning against the wall of my garage, 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: it represents 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 subjects 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 crucially for this essay, it is reflected in relations of direction: thus 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 function of the visual system that describe it as reconstructing a three-dimensional model of the external world given the retinal input.
A detailed version of the metric visual space idea is developed by Peacocke (1992). 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.

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.

3 Another careful development of the metrical space account can be found in Matthen (2005, pp. 271-289; 2014, pp. 266-279).

4 Officially, Peacocke’s notion of SCENARIO CONTENT is a set of scenarios, not a singleton scenario. But he indicates that the introduction of sets is intended only to model variations in perceptual acuity, the degree of clarity or resolution in a perceptual state, like that exhibited by the differences between focal and peripheral vision (Peacocke 1992, p. 63). (My own treatment of visual acuity, discussed in Section 5, derives instead from the granularity of cells in the visual field.) 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. But if the set of scenarios in scenario content were allowed to vary with respect to any kind of feature, and not just that of acuity, then Peacocke’s view would have more in common with centered-worlds approaches to content discussed in the next section. Then depth-indeterminacy would not be a problem, but the theory would still face the challenge of contradictory content.

§2 Visual content as metric space
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.

Indeterminacy with respect to depth typically implies indeterminacy about size as well, since if a given object (like a distant pyramid) is closer, it must be smaller, or if further away, then larger. Depth indeterminacy can also give rise to shape indeterminacy, since 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. Thus even relatively rich, mid-level visual perception seems to be capable of indeterminacy about the metric properties of depth, size, and shape. These facts directly contradict the idea that visual space is metric.

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.

We can now see a kind of dilemma for any account of visual content. On one hand, it must do justice to the obvious sense in which visual representations express unified spaces of some kind. On the other hand, the relevant notion of space must allow for indeterminacies of depth, size, and shape. 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 (or centered-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. (I’ll elaborate on the relevant notion of viewpoint in Section 6.) 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 magni-
tude 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 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, at right. The situation it depicts seems to be impossible, for it shows the same stream of water continually flowing downwards, but ending up in the same place it began. Only here, the drawing clearly depicts its visual contradiction in the context of an otherwise coherent visual space. The problem at hand isn’t exactly that the

Figure 2: Impossible images: at left, the Penrose triangle (Penrose and Penrose 1958); at right, Escher’s Waterfall (1961).
unstructured view assigns this picture *no* content, but rather that it assigns the *same* null content to all contradictory pictures, no matter how different the spaces they express. The challenge posed by contradictory images to the unstructured view is that it fails to make distinctions among contents where they ought to be made. Thus, at least for contradictory pictorial contents, the unstructured view falters.

Can the same point be extended to perceptual content? Arguably, contradictory perceptual content is generated by simply *looking* at drawings like the ones shown here (Peacocke 1992, p. 74). But the case isn’t clear cut: it’s possible that the visual system cannot represent the entire content of an Escher drawing all at once, but only partial, consistent fragments, giving rise only to fragmentary but non-contradictory perceptual content.\(^5\)

But there are other cases which more clearly give rise to perceptual contradictions. The so-called “waterfall illusion” (unrelated to the Escher waterfall above) is one: 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). Other cases may be even closer to home. 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; “this double reality is part of the paradox of pictures” (Gregory 1970, p. 22).\(^6\)

Still further cases may involve intentional identities and non-identities (Blumson 2009, §5; Millar 2016). Two objects, presented in different locations on the visual field, or different points in time, might be perceived as one. Or a single object, presented at different locations on the visual field or in time, might be perceived as two. In the latter case, for example, two ends of a snake whose mid-section is occluded might be perceived as disconnected, hence perceived as belonging to distinct objects. Yet unstructured content is a poor candidate for the content of such a state, for there are no worlds where the very same object is also two objects. A careful description of cases like this, and of the representational mechanisms involved, would take us too far afield. At the same time, such cases are, at the very least, in-principle possible, and our conception of visual content should be broad enough to incorporate them.\(^7\)

---

\(^5\)There is some empirical evidence for this claim, see for example, Schacter et al. 1991. But it is difficult to distinguish here between the ability of the perceptual system to *represent* impossible contents, and its ability to *preserve* such representations in memory.

\(^6\)In a future draft, I hope to expand this section to respond to the worry that the perceptual cases surveyed here are in fact examples of bi-stability, not contradiction. One possibility is that, even if short term saccade-based perceptual representations are bi-stable, visual representations which integrate these views into longer term memory may be more clearly contradictory.

\(^7\)A complication is that, in the actual working of the perceptual system, such identities seem to arise chiefly from operations of co-indexing, rather than explicit statements of identity (Pylyshyn 2007). But then it is not clear whether false identifications may arise, since co-indexed representations which bear the relevant causal connections to more than one
Such phenomena appear to pose a basic challenge to the unstructured approach. Perhaps some headway can be made 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 representation 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.\footnote{As far as I know, only 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. Such a sequence might include, for example, a given object, the property circular and the property square. My own proposal ultimately makes use of some of the same ingredients. But Blumson's strategy crucially loses the idea that visual representations express an entire space of objects and properties, and not just a single basic proposition—a feature preserved in the initial centered-worlds proposal. What is really called for is a layer of additional structure, more than the unstructured view affords, but one that also provides the perspectival spatial unity distinctive of visual content.}

As far as I know, only 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. Such a sequence might include, for example, a given object, the property circular and the property square. My own proposal ultimately makes use of some of the same ingredients. But Blumson’s strategy crucially loses the idea that visual representations express an entire space of objects and properties, and not just a single basic proposition—a feature preserved in the initial centered-worlds proposal. What is really called for is a layer of additional structure, more than the unstructured view affords, but one that also provides the perspectival spatial unity distinctive of visual content.

In fact, 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” (similar to Blumson’s sequences above) 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.

In fact, 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” (similar to Blumson’s sequences above) 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.

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 turned 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.

In addition, Peacocke draws attention to further phenomena which cause equal trouble for both metric visual space views (minus the protopropositions) and unstructured views of visual content. These are perceptual attributes which are clearly represented, and phenomenologically

\begin{itemize}
  \item object likely do not refer at all. Conversely, false distinctions may arise, but only if the system has a mechanism to reliably signal non-identity; mere non-co-indexication does not typically imply non-identity in familiar linguistic systems.
  \item Another problem for the fragmentation approach, suggested to me by Justin D’Ambrosio, is that there are, in general, innumerably many ways of dividing a given visual content into consistent fragments, thus saddling the fragmentation theorist with an arbitrary selection in nearly every case.
\end{itemize}
accessible, but do not make a difference to accuracy conditions or the spaces filled out. Peacocke (1992, pp. 75-76) uses the vivid example of shape orientation: in Figure 3 below, the image at far left is perceived as a diamond, while that in the center as a square. The difference cannot be accounted for merely in terms of the relationship between the image and the orientation of the viewpoint, because the interior shape at far right is tilted, but still perceived as a square. (Other visual phenomena, such a perceptual grouping, would make the point just as well.)

![Figure 3: An apparent diamond, square, and square inside a rectangle. (The first two images are from Peacocke 1992, the third is from Palmer 1983.)](image)

The distinction between a diamond and a tilted square makes no difference for accuracy or the way space is filled out, so cannot be encoded in the corresponding conceptions of content. The lesson is much the same as before: there should be a way for visual contents to make piecemeal attributions of features at a finer level of granularity than that of concrete or metric spaces. While it is natural to think that some kind of structured object, akin to a structured proposition, will satisfy this demand, we should not rush to equate visual contents with structured propositions, for we must also maintain the sense in which visual representations express perspectival visual spaces. This is the project of the next section.

## 4 Perspectival feature maps

### 4.1 The proposal

In this section I outline an alternative account of visual content. I propose to hold fixed the directional organization of space, building it into the structure of visual content, while eschewing the structural commitment to depth of the metric space theory. Meanwhile, attributes of depth, along with color, shape, and other features, are are included piecemeal, making room for the expression of indeterminate and contradictory visual content. Finally, I propose to unify these elements within a coordinated visual frame, the semantic reflection of the visual field. The result is a kind of structure I'll call a PERSPECTIVAL FEATURE MAP. My claim is that all visual contents are perspectival feature maps. While different visual modalities fill out perspectival feature maps in different
ways, all express contents of the same basic form.

I begin by briefly outlining the structure and intended interpretation of perspectival feature maps; I develop these themes in more detail in Sections 5 and 6, and in the appendix. Perspectival feature maps are types of abstract object. They consist of a core geometric structure, essentially similar from case to case, and set of variable elements that include particular objects, properties and relations. 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 points 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.

![Diagram of map field and directions](image)

**Figure 4:** The core structure of a perspectival feature map: the map field and its directions.

The variable elements of a perspectival feature map are embedded in what I call FEATURE CLUSTERS. Each feature cluster contains an object and set of properties. Relations are linked to sequences of feature clusters. So while perspectival feature maps themselves are abstract structures, they have concrete objects, as well as properties and relations as constituents. In this respect, they are comparable Russellian structured propositions, which are themselves abstract objects in which concrete objects as well as properties and relations inhabit the nodes of a syntactic tree. Metaphorically, a perspectival feature map is like a drawing made from bits of reality—constructed not from lines but actual edges, not from images but actual objects.

§4 *Perspectival feature maps*
In a perspectival feature map, points in the map field are associated both with directions, part of the core structure, and clusters of objects and properties, part of the variable structure. The feature map itself is a kind of directional space, locating each such object and property in its associated direction. It is accurate at some scene when, for every feature cluster in the map, the object in that feature cluster is located in the direction given by the map, and instantiates the features it is linked to.\(^9\)

This view understands visual content to embody a certain kind of array-like structure which may be contrasted with the structured propositions traditionally associated with sentences and thoughts. The constituency structures of such propositions are tree-like, not array-like, and each leaf of the tree is associated with a single object, property, or relation, not a cluster. Perspectival feature maps themselves are nominally “propositional” in the sense that they are object-involving structures which determine accuracy conditions, or sets of centered-worlds. But they are not propositional in the sense that they are not the tree-like structures of Russellian propositions (Byrne 2001, pp. 201-2; Crane 2009; Grzankowski 2015; Camp 2018).

Although propositional structures are the familiar determiners of truth, there is no reason to think that perspectival feature maps are too exotic to fix conditions of accuracy. After all, propositional structure is just one kind of structure. The primitive structural relation in a tree-like proposition is (depending on specifics) something like a triadic relation between ordered sibling nodes and their parents. In perspectival feature maps, the building blocks are the geometrical relations

---

\(^9\)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.” Though Matthen does not go on to elaborate his vision of the “omnipresent grid.”
that define the map field and its directions. Perspectival feature maps simply express contents whose organizations draw from a different structural domain.

4.2 Visual field and visual direction

The core structure of a perspectival feature map is comprised of two elements: (i) the two-dimensional map field, the abstract counterpart of the visual field; and (ii) 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 associated with a region of the visual field. And second, every such object and property is located along a viewpoint-centered direction. Although different representational systems exploit different kinds of map field and directional organization, I'll focus here on flat rectilinear map fields, like that shown above, for reasons of exposition. Variations are discussed in Section 5.

The primary role of the map field is to constrain the scope of visual content. In particular, visual representations do not normally locate objects in all directions from the viewpoint, but only within a predefined window of a particular shape. The map field encodes this shape as a structural limit on content.

A similar consideration has to do with the restriction of spatial scope in the forward-backward dimension. Visual content represents only those objects which lie in front of the visual field. Objects which are within the directional window of the map field, but are nevertheless behind the map field, are not represented. The significance of this fact is negligible in perception, where the only such objects are microscopic particles inside the eyeball. But in pictorial representation, especially depiction with a deep focal length, there may be a whole range of objects between the projection source and the map field which are simply omitted.

Finally, as discussed in Section 5 map fields are divided into cells, providing visual content with a base-line level of acuity or resolution. It's a given that different visual representations have different levels of acuity (perhaps in different regions of the visual field), and this is modeled in part by the size, shape, and distribution of cells on the map field. They determine a minimal “chunking” of perspectival directions into bundles, itself reflected in the variable granularity of visual content.

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. More carefully, visual contents are always determinate with respect to direction up to the level of acuity allowed by the map field in question. 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. Direction is not like depth in this respect, which may vary
across the visual field from representation to representation.

There has been some debate in perceptual science as to whether low-level perceptual representations invariably locate their attributes in directional space (or on the visual field). Early findings by Treisman and Schmidt (1982), for instance, show that perceived features often float free from their retinal positions; a possible interpretation was that direction (like depth) is a common but inessential aspect of perceptual content. But later work has confirmed that, though the perceptual system seems to make errors about attribution of direction, there are no clear examples in which it is not represented at all (Ashby et al. 1996; Pylyshyn 2006, pp. 177-8).

While visual field and visual direction are fixed aspects of the structure of perspectival feature maps, the elements of the feature clusters may vary from representation to representation. In particular, depth is treated like just another optional feature. An object may be represented as red or square, or as lying at a given distance from the map field. Further, the representation of depth may be absolute in some cases, indeterminate in others, and wholly absent in still others. This is because the feature map may include an absolute depth property like *five feet from the viewpoint*, but it may also contain only indeterminate depth relations like *further from the viewpoint than*. Unlike metric visual spaces, relations of depth aren’t built into the structure of perspectival feature maps. In a similar fashion, because represented properties are allowed to combine freely, unconstrained by spatial structure beyond direction, spatial inconsistencies are easily captured. Both cases are discussed in more detail in the next section.

Having described the core structure of perspectival feature maps, we should next ask what justifies the attribution of this common structure to the varied domains of visual representation. Why should we expect this account of visual content to be right? A number of factors motivate my approach here.

First and foremost, 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 associated with them. Perspectival feature maps encode what is essential to visual content in their core structure, relegating what is contingent to the feature clusters. A complication here is that different systems of representation recruit different kinds of visual fields and different organizations of perspectival directions, as I’ll discuss in Section 5. More carefully, then, the core structure of a perspectival feature map corresponds to those aspects of visual content which are fixed across visual representations within a given system.\footnote{Matthen (2005, pp. 274-276) defends such an asymmetrical treatment of direction and depth. He observes that the same direction relations are present in every perceptual visual episode, but not the same depth relations. Relatedly, each direction relation appears only once in each perceptual visual episode, but the same depth relation may appear multiple times. As Matthen writes on p. 276, “distance is beginning to seem more and more analogous to colour, shape, or motion.”}

\section{Perspectival feature maps}

\footnote{The individuation of systems is further complicated by the possibility that some systems may include adjustable}
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 the incoming light to the retina. 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.

Finally, the distinction between core and variable structure seems to reflect different ways that content is encoded in the course of visual computation. Features like color, shape, or motion are thought to be explicitly signaled, 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, rather than the explicit output of a computation. Content implicitly encoded in the fixed causal structure of the perceptual system is reflected in the perspectival feature map’s core structure. While explicitly encoded content, such as perceptual representations of objects or features, 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 worlds 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 Feature maps in perceptual science

Perspectival feature maps are natural candidates for the contents of visual representations, for they recapitulate the retinotopic structure of visual computation itself. The use of feature maps to analyze perceptual states dates back at least to the work of Marr (1982) and Treisman (1988), and is now ubiquitous in vision science and computer vision. Feature maps are invoked both as data structures in algorithmic descriptions of visual processing, and as physical structures at the level of neural implementation. Starting with the retinal image, the visual system is thought to process information registered in successive retinotopic brain maps, embellishing the information distributed across discrete regions of the image at each pass (Frisby and Stone 2010, ch. 10).

parameters for focal length and visual field shape. Still, in perception, core structure is fixed by endogenous manipulations of visual system (including, e.g. the size of the attentional window or the focal length of the eye), while variable content is determined by the response of perceptual algorithms to the variable retinal input.

12Thanks to William Kowalsky for help with this point. See also Matthen (2005, p. 276).

13The terminology of “feature maps” comes from the work of Treisman (1980; 1986; 1988). She uses the term to refer to the representation of a single determinate feature across the visual field. By contrast, feature maps in my sense can contain a range of different features at once, and may bind these features together with objects. Treisman’s feature maps (qua representations) do express feature maps in my sense (qua content), but they are not the only thing that does.
A typical use of feature maps from the literature, in this case from Marr (1982, p. 278), shows a distribution of orientation features across a visual field. The arrows here are visualizations of symbolic representations of surface orientation (not direction from the viewpoint!).

![Figure 6: A typical feature map from Marr (1982, p. 278): arrows represent local surface orientations, solid lines represent occlusion contours; dotted lines represent surface orientation discontinuities.](image)

The present proposal departs from common trends in this literature, exemplified by Marr’s use here, in several ways. First, feature maps are typically deployed in the analysis of a specific visual modality, such as mental imagery, or a particular subsystem of vision. The aim of the perspectival feature map account is to identify that relatively sparse core structure which characterizes content across visual modalities.

Second, nearly all uses of feature maps in perceptual science, including the standard case displayed above, assign an explicit role only to the features they attribute to the environment, making no explicit allowance for the representation of individual objects. Perspectival feature maps bring this important feature of visual content to the fore, and showing concretely how singular content combines with features and directions to determine precise accuracy conditions.

Third, feature maps are typically construed as descriptions of representational vehicles, information structures at the algorithmic or implementation levels of Marr’s explanatory hierarchy (but see e.g. Haugeland 1991 or Burge 2005, p. 3). Consequently, the cells of feature maps are typically associated with symbols, or equivalent representations of features. By contrast, the contents of a perspectival feature map are actual object and properties, not representational vehicles. In this respect, they are comparable to structured propositions.

Finally, what is most distinctive about perspectival feature maps is the explicit inclusion of the directional array as part of the core structure of the feature map. Most 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)). They do not make manifest how or whether directions figure into content or conditions of accuracy. In the theory of perspectival feature maps, by contrast, the role of perspectival direction is made explicit, built directly into the structure of the feature map itself, and embedded in the accompanying definition of accuracy.

4.4 Mark of the visual

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, even though they exhibit spatial structure and engage visual cognition. This is 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 mental map, for example, might represent the location of various landmarks within a fixed coordinate system (Camp 2007, p. 158). But because 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. For example, 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: \textit{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. They only become parts of a visual representation when they are embedded in a viewpoint-centered description of space. Thus visual contents seem to form a natural class within the more general class of spatial contents.

For similar reasons, not all representations computed on the basis of optical information are necessarily visual. Suppose an animal is discovered which has a rudimentary perceptual system that is based on a light-sensitive retina, but outputs representations which do not preserve directional information in any way. Perhaps, on the basis of a distinctive optical cue, some module outputs a representation with the content \textit{there is a dog nearby}, but assigns no directional position to the dog, not even restricting it to the field of view. In that case, I am inclined to insist that the representation in question is not in fact a \textit{visual} representation, despite the fact that it is the result of processing a complex optical signal.

Ultimately, perspectival feature maps are natural tools to understand visual content because
they bear a direct relation both to the structure of visual representations themselves, and to the
spaces those representations express. The two-dimensional map field in a perspectival feature
map is what gives the account its “pictorial” flavor; while the array of three-dimensional directions
is what ensures that the account is still a theory of “visual space.” Perspectival feature maps offer
an accessible handle by which to grasp both the two- and three-dimensional aspects of visual
representation at once.

5 The structure of perspectival feature maps

In this section I develop the definition of perspectival feature maps initially sketched above,
and in the next section, I describe their accuracy conditions. A formalization of these ideas is
provided in the appendix.

Every feature map is built upon a BASIC MAP FIELD (for short, a MAP FIELD), the abstract
counterpart of the visual field or picture plane. The central component of a map field is a segment
of a two-dimensional surface. I’ll typically discuss planar map fields, but they may also be curved,
and this may be the appropriate choice for perceptual content. The map field is embedded in a
three-dimensional Euclidean space, what I’ll call the space of the map field— not to be confused
with the “visual space” embodied by the feature map itself.

The map field surface is coupled with a pair of orientation vectors— one, parallel to the plane,
indicates the map field’s up-down orientation; the other, perpendicular to the plane, indicates its
front-back orientation. The shape of the map field loosely reflects the organization of the visual
field of a given visual representation. Thus, for a rectangular painting, the feature map will be
based on a rectangular map field. While for perception, it will have the shape of the perceptual
visual field, a kind of bent oval.

The map field is divided into an array of cells. The number of cells may vary, corresponding
loosely to the acuity or resolution of the type of visual representation in question. There may be
finitely many cells, as for standard cases of digital depiction, or continuum many, perhaps, for
analogue pictures or vector graphics. In addition, the arrangement of cells will vary depending
on the underlying system of representation. Human visual perception is plausibly represented
by a large but finite number of partially overlapping cells; the organization of cells may directly
reflect the structure of receptive fields in the visual system, or it may derive from facts about acuity
at higher levels of abstraction. The cells would likely be more densely packed in the center than
the periphery, corresponding loosely to the greater focal acuity (and greater number of receptive
fields) in the central regions of mammalian vision. For a digital photograph, by contrast, they
would be distributed in an even, non-overlapping grid across the map field. See Figure 7 for

such cells need not form a partition, for they may overlap, and in principle they need not cover the entire field. For
purposes of illustration I will typically treat them as forming a partition.
In what follows, I’ll illustrate my discussion with flat, square-shaped map fields, uniformly filled with finitely many non-overlapping cells. My comments about these simple models should extend, with some amendments, to the other, more complex and biological cases.

![Figure 7: At left, a uniform map field. At right, the graduated map field of monocular human vision (Ruch and Fulton 1960).](image)

A **feature map** is a map field in which some regions are populated with objects and properties, as illustrated in Figure 8. I will call such object-laden regions **segments**. Typically, every cell constitutes a segment, and I’ll ultimately allow for multi-cell segments as well. I assume a primitive structural relation of **linking** which associates segments with **feature clusters**. A feature cluster has two parts, modeled as an ordered pair. The first element of the cluster is a particular object, while the second element is a set of properties or **features**. The objects in feature clusters correspond to a visual representation’s singular content, while the properties in feature clusters correspond to its attributive content.

---

15 The basic idea of a feature cluster is prefigured in much work on visual content, beginning with Treisman’s (1980) proposals that visual features are integrated into a single object representation. Other authors have employed counterpart notions, including the “proto-propositions” of Peacocke (1992), the “symbol vectors” of Tye (2000), the “feature-placing structures” of Matthen (2014), the “multiple-slot memory” of Green and Quilty-Dunn (2017), and the “noun-phrase structures” of Burge (2018). My conception of a feature cluster aims to crystalize this tradition.
In this essay, I consider only cases of visual content that include reference to particular individuals. But a variety of important phenomena are thereby neglected. These include indefinite content, as when a picture depicts some cube rather than a particular cube; discourse reference, as when a picture depicts the same cube as was depicted by another picture (or described in another sentence); hyperintensional content, as arises in the deployment of visual indexes in perception; and the content of pictures with failed reference, as in the case of hallucination (Burge 2010; Schellenberg 2018). All these phenomena suggest that something like Frege’s notion of an object sense be used in place of concrete objects in the specification of visual content. Similar forces pressure one to adopt a Fregean, as opposed to Russellian, view of structured propositions. My intention for now is that the objects I enlist to characterize feature clusters can ultimately be substituted with object senses, discourse referents, or the like, in order to model a wider array of phenomena. This is one of the ways the feature map idea provides a framework within which to work out open questions in philosophy of perception. For ease of exposition, I’ll stick with concrete individuals here.

Thus far I have written about feature maps as if only individual cells are selected to be content-bearing segments. But this restriction must be relaxed. Perception is not limited to representing the locations and features of millions of tiny surfaces and edges—though it does seem to do this too. In addition it represents the locations and features of larger objects, whose images extend across the visual field. The way to handle this is straightforward: in a feature map, feature clusters are associated not only with individual cells, but with sets of cells as well. Not every set of cells, of course, will be counted as a segment in this way; only those sets of contiguous cells which correspond to the visual image of an object will become a segment and assigned a feature cluster. In the resulting feature map, many feature clusters will be associated with overlapping segments.
of the array, but this need not cause any difficulty.\footnote{One may be tempted to assume that if a set of cells $A$ is a subset of a set $B$, and both are associated with feature clusters, then the $A$-object will be a part of the $B$-object. But as I will discuss in Section 9, the phenomena of transparency belies this appealing hypothesis in its full generality.}

![Figure 9: Basic feature map with feature clusters at different scales.](image)

The discussion so far has also focussed exclusively on the attribution of visual \textit{properties}, but we must also allow for visual \textit{relations}. Candidates include relations between objects like \textit{being the same size as} or \textit{being a darker color than}. Such relations cannot be structurally located within feature clusters, because feature clusters are associated with single objects, while relations hold between objects. Instead, I will think of relations as linking feature clusters together. (The whole structure now looks a bit like streamers hanging from ceiling tiles.) In this context, the concept of \textit{linking} is extended to describe a primitive structural relation that connects a sequence of feature clusters to a relation. An example is illustrated in Figure 10.

![Figure 10: Basic feature map with relations.](image)

How should we interpret a feature map? The guiding intuition is that a feature map is accurate

\section{The structure of perspectival feature maps}
when: (i) the world contains the objects in each of its feature clusters; (ii) each of those objects instantiate the properties associated with them; and (iii) the various objects of the feature clusters bear their associated relations to one another. What this description leaves out is any sense of the spatial distribution implicit in the dimensions of the map field. For all I have said so far, the simple propositions associated with each segment of a feature map need not have any particular spatial significance or unity. But any apt analysis of visual content must capture some basic aspect of visual space, without, of course, committing to metric visual space. I propose that the key feature of the space defined by a feature map is that it is \textit{perspectival}, a kind of directional space centered at a viewpoint.

To capture this idea, I treat the concept of \textit{direction} as primitive. Directions indicate orientations or linear paths; a direction is the residue of a vector when magnitude is removed. Direction in this sense is to be contrasted with distance, which specifies how far an object is from a reference point, but not which way it is relative to that point. In this essay, directions are always defined relative to a definite starting point in space, the point from which they indicate an orientation. Later I’ll formally model directions as \textit{rays}, half-lines which have an initial point, but extend infinitely away from it. Directions can also be compactly represented within a coordinate system, either as unit vectors or in terms of angles in three-dimensions.

A \textit{directional map field}, like that shown in Figure 11, is a basic map field for which each point in the field is associated with a direction. All such directions point towards the front of the map field (but are not necessarily parallel with its front-back orientation). Each \textit{point} on the field can be thought of as pointing in a \textit{particular} direction, and each \textit{set} of points indicates a \textit{bundle} of directions. Thus each cell and segment of the field will be associated with bundles of directions, and these in turn are used to locate feature clusters within visual space.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{angular_map_field.png}
\caption{An arbitrary directional map field and a perspectival map field.}
\end{figure}
A **perspectival map field** is a directional map field where the directions are organized as if around a unified visual perspective. The directional unity of a perspectival map field is defined in terms of projection from a viewpoint. To see this, consider a map field in the context of a three-dimensional space, as in Figure 12. Suppose a geometric point, which I’ll call the **projection source**, were located at some distance from the field of the feature map.\(^{17}\) Then lines could be drawn from the projection source through each point in the visual field; I’ll call these **projection lines**. The path of each projection line from the projection source through the map field, and continuing on beyond it, defines a direction. A perspectival map field is any directional map field that *could* be embedded in a three-dimensional space in relation to a projection source such that the directions of the map field would be collinear with the resulting projection lines. Since projection sources are not in fact part of visual content itself, the directions in a perspectival map field are officially defined relative to points in field’s surface. But since all such directions are required to converge backwards on a hypothetical projection source, I’ll say that they are collectively **centered** at that source. In fact, I’ll typically say that directions are centered at a *viewpoint*, a notion which I’ll precisify in the next section.

![Figure 12: Projection lines from a point source to a basic map field, left, define a perspective map field, right.](image)

The next step is to combine a perspectival map field with an arrangement of feature clusters. In the resulting **perspectival feature map**, each feature cluster, in virtue of its link to a segment of the map field, is also associated with that segment’s bundle of directions. In effect, the perspectival feature map locates each of its feature clusters in its own proper direction, as in

\(^{17}\)Typically the projection source would be located behind the map field; systems of inverse projection require an exception to this rule.

§5 *The structure of perspectival feature maps*
Figure 13: Perspectival feature map.

A perspectival feature map can be thought of as a kind of directional space—a space whose “dimensions,” speaking loosely, are directions emanating from a viewpoint. This space in turn defines angular directions, not just up-down, and left-right, but quantitatively distinct directions for each cell and segment in the map. If one were to define a space using a polar coordinate system and then remove any representation of distance, you’d be left with a description of directional space in many ways like that embodied by a perspectival feature map.

In fact, the arrangement of directions as converging backwards on a point describes just one possible directional space among many. Pictures created in systems of parallel projection, like the axonometric and oblique projections often used in technical drawings, give rise to different types of directional space, as in Figure 14. To define the perspectival map fields for representations in such systems, we employ alternative projection sources. For images in parallel projection, the projection source is a plane; all projection lines are normal to the plane, and they intersect points in the map field. As before, the directions of the map field must be aligned with these projection lines. For some types of parallel projection the projection source (now a plane) is parallel to the map field, in others, it is positioned at an angle to it. Either way, parallel systems give rise to pictorial contents with a kind of unsituated “god’s eye view.” But such contents are still perspectival, since they clearly represent their subjects from a particular direction, and I will still refer to the corresponding perspectival feature maps as centered at a viewpoint.

§5 The structure of perspectival feature maps
Figure 14: Perspectival map fields based on two different parallel projection systems.

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

Capturing the content of such states using perspectival feature maps requires that we distinguish between 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. Determinate depth properties described above are a prime example; attribution of depth effectively describes an object as being some 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. In general, perspectival properties differ from other kinds of features because they are defined in relation to positions in the visual field, rather than exclusively in terms of the intrinsic characteristics of the objects in question.

Besides perspectival properties, there are also perspectival relations—relations between represented objects relative to a viewpoint. For example, the relation more distant than holds in the first place between two objects, but makes implicit reference to the viewpoint. The inclusion of such relations is straightforward and inevitable in the current framework.

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.\textsuperscript{18} 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 15; 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.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{feature_map.png}
\caption{Indeterminate depth relations in a feature map.}
\end{figure}

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 edge, relative to the viewpoint. Labeling the three facing surfaces in the figure $O_1$, $O_2$, and $O_3$,\textsuperscript{18}

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.

\section*{5 The structure of perspectival feature maps}
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 brings out, 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 16: Inconsistent spatial relations in a feature map.](image)

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

### 6 Accuracy for perspectival feature maps

I’ve claimed that visual contents are perspectival feature maps. For a theory of content to earn its name, however, it must specify a clear mechanism by which its chosen conception of content determine conditions of accuracy or satisfaction.\(^{19}\) In this section I provide such a definition for perspectival feature maps.\(^{20}\)

\(^{19}\)Here I face the problem of the unity of the proposition, as applied to perspectival feature maps. I assume that the structure of perspectival feature maps and the corresponding definition of accuracy are intimately connected in some way. In a future draft, I’ll discuss how different solutions to the problem of unity can be applied to the case at hand.

\(^{20}\)In the framework I employ here, a visual content is accurate or inaccurate relative to an index. In an alternative framework (in fact, the one I use in Greenberg 2018a), it is representations which are accurate or inaccurate, while contents
Intuitively, 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. It thereby provides a kind of direction-based description of a the world beyond the visual field. It is accurate in some situation when that description is satisfied.

To develop this idea in detail, we distinguish a perspectival feature map from the index situation relative to which it accurate or inaccurate. My guiding assumption is that a visual content is accurate only relative to a VIEWPOINT-CENTERED WORLD, the pair of a world and a viewpoint, which I’ll call a SCENE. In effect, the world supplies the facts which the feature map is compared to, and the viewpoint supplies the perspective on those facts which the feature map purports to embody. Feature maps are abstract, partial descriptions of space, unbound to any particular possible world. By contrast, scenes are concrete, complete, and determinate in every respect, built from a single possible world. As a consequence, feature maps are always realized by multiple scenes.

In this context, VIEWPOINT refers to a particular oriented location in space in a particular possible world. A viewpoint need not be tied to any real, implied, or even imagined viewer. It is simply a spatial index.

The viewpoint-relativity of accuracy is motivated by the observation that the content of a given percept or picture may be accurate at some viewpoints, but not others— even relative to the same possible world. The content of my current perceptual state— the desk here, the computer there, and so on— is perfectly accurate relative to my own vantage point; but it would be inaccurate relative to a vantage point just a foot to the left. For although nothing in the scene itself would be different, the represented perspectival relations of direction, depth, and shape would no longer accurately describe the scene.

I refine the definition of viewpoint using the ingredients of geometrical projection (Greenberg 2018b). A viewpoint is composed of a pair of indices, the first of which gives the location of a projection source, and the second the spatio-temporal location of a projection plane. For viewpoints in linear perspective, the projection source is simply a point in space. The projection plane is a segment of a plane, supplemented with up-down and front-back orientation vectors. This plane provides the “window” or “viewfinder” of the viewpoint. Together these elements make up what I have called an “oriented location,” where the projection source fixes the location of the viewpoint, and the position of the picture plane relative to the projection source determines its orientation. This idea is brought out by visualizing the bundle of projection lines that start at the projection source and either hold or fail to hold at an index. The difference is largely terminological. The key point is that contents determine conditions in which they are satisfied, whether satisfaction is called “accuracy” or something else.

21Officially, viewpoints also specify a time within the relevant world; I abstract from time throughout this essay, save for the comments in Section 9.7.

22An alternative view holds that visual content is viewpoint-indexical, i.e. that particular concrete viewpoints are built into the content itself. A variety of reasons lead me to reject this idea, but the issue is too complex to adjudicate here. Still, proponents of viewpoint-indexicality should have little trouble translating my proposal into their preferred idiom.
source and pass through the projection plane.

![Figure 17: A viewpoint consists of a projection source and projection plane.](image)

The evaluation of a perspectival feature map relative to a viewpoint-centered world takes place in two stages. The feature map itself is an abstract, geometrical object, defined within its own three-dimensional Euclidean space. But a centered-world is made up of a possible world, understood as a concrete way things could be, and a viewpoint located within the three-dimensional space of that world. To compare the feature map to the centered-world, we first embed the map at the location of the viewpoint; then, in order to derive an accuracy value, we may evaluate whether the constraints the feature map represents hold relative to the location where it is embedded. I develop this strategy in a series of steps.\(^{23}\)

First, I refine the informal idea of embedding a feature map in a world using the concept of CONGRUENCE. A perspectival feature map is congruent with a viewpoint when there is a metric isomorphism between the perspectival feature map and the viewpoint such that three constraints are satisfied. First, the map field must be the same shape and size as the projection plane. Second, the map field must have the same front-back and left-right orientation as the projection plane. And third, the array of directions built into the map field must match the directions defined by the projection lines within the viewpoint. This amounts to the first clause of the definition of accuracy:

\[
\text{A perspectival feature map } M \text{ is accurate at centered-world } \langle w, v \rangle \text{ only if } M \text{ is congruent with } v.
\]

In this context, I’ll adopt the convention of referring to elements of the map field with unadorned variables (direction \(d\), point \(p\)), and referring to their isomorphic counterparts within the space defined by the congruent viewpoint with \(v\)-subscripts (direction \(d_v\), point \(p_v\)).

\(^{23}\)Peacocke (1992, p. 64) develops a definition of correctness for scenario content along similar lines.
Next, I use the directions built into the perspectival feature map to specify the locations of the objects and properties in the feature clusters. Let’s say that a segment $S$ is linked to a feature cluster $C = \langle O, F^* \rangle$, where $O$ is an object and $F^*$ is a set of properties, and also that $S$ is associated with direction $d$. Then the feature map will be accurate only if $O$ is located along $d_v$. This condition, however, is too rough: in the general case, segments are not associated with individual directions, but with bundles of neighboring directions. The way to handle this is typically to require that some part of the object in question be located at every direction in the bundle. So letting $D$ be the bundle of directions associated with $S$, the feature map will be accurate only if, for every $d_v$ in $D_v$, some part of $O$ is located along $d_v$.

I add this to definition:

for every segment $S$ in $M$ and feature cluster $C = \langle O, F^* \rangle$ linked to $S$:

where $D$ is the bundle of directions associated with $S$:

for every direction $d_v$ in $D_v$: some part of $O$ is located along $d_v$.

A separate clause ensures that the properties in each feature cluster are instantiated by the object of that feature cluster. I divide this constraint into two parts, one devoted to non-relational properties (like being spherical), and one to perspectival properties (like being tilted away).

for every non-perspectival property $F$ in $F^*$: $O$ is $F$ at $w$;
for every perspectival property $F$ in $F^*$: $O$ is $F$ relative to $v$ at $w$.

The last ingredient is the treatment of relations. Here again, I introduce separate clauses for perspectival and non-perspectival relations. And I adopt the convention of referring to the object of feature cluster $C_n$ as $O_n$.

for every sequence of feature clusters $C_1, ..., C_n$ linked to segments in $M$:

for every non-perspectival relation $R$ linked to $C_1, ..., C_n$: $O_1, ..., O_n$ stand in $R$ in $w$;
for every perspectival relation $R$ linked to $C_1, ..., C_n$: $O_1, ..., O_n$ stand in $R$ to $v$ in $w$.

I finally combine the four conditions reviewed above: the constraints on congruence, object direction, property instantiation, and relations. The result is a set of necessary and sufficient conditions on accuracy for perspectival feature maps.

A perspectival feature map $M$ is accurate at centered-world $\langle w, v \rangle$ iff

(i) $M$ is congruent with $v$;
for every segment $S$ in $M$, and for every feature cluster $C = \langle O, F^* \rangle$ linked to $S$:

(ii) where $D$ is the bundle of directions associated with $S$,
for every direction $d_v$ in $D_v$: some part of $O$ is located along $d_v$.

---

Certain kinds of feature maps may require variation in the quantifier here. I have in mind, for example, Those cells which correspond to the pixels of a digital photograph. A colored pixel plausibly describes the world as containing an object which is intersected by some or most of the directions in its bundle, but not all.

§6 Accuracy for perspectival feature maps
(iii) for every non-perspectival property \( F \) in \( F^* \): \( O \) is \( F \) at \( w \);
for every perspectival property \( F \) in \( F^* \): \( O \) is \( F \) relative to \( v \) at \( w \);
for every sequence of feature clusters \( C_1, ..., C_n \) linked to segments in \( M \):
(iv) for every non-perspectival relation \( R \) linked to \( C_1, ..., C_n \): \( O_1, ..., O_n \) stand in \( R \) in \( w \);
for every perspectival relation \( R \) linked to \( C_1, ..., C_n \): \( O_1, ..., O_n \) stand in \( R \) to \( v \) in \( w \).

This definition reveals the sense in which perspectival feature maps qualify as content. For we can now see that perspectival feature maps are not merely structures, they are representational structures with conditions of accuracy.

7 Further conditions on visual content

I’ve suggested that expressing a perspectival feature map as content is an essential condition on what it is for a representation to be visual. But is this condition also sufficient? This is a more difficult question, which I do not intend to settle here. One might, for example, look to the type of causal conditions or mechanisms that produced the representations in question in the first place. But even if we stay focussed on content, I wish to flag two areas where theorists might reasonably look for additional necessary conditions on visual content.

First, the present proposal puts few constraints on what kinds of properties and relations are expressed by visual representations, save that they be instantiated by the kind of objects that can occupy spatial locations. But it is certainly conceivable that only some properties which are compatible with these constraints are properly visual. One might think, for example, that visual representations must at least represent properties of surfaces, or perhaps that they must represent visible features or spatial features. These suggestions are plausible, reflecting as they do core cases of perceptual representation. But there are also apparent counterexamples. Wire-frame drawings are clearly visual, but do not obviously represent surfaces, and certainly not visible surfaces. Heat-map photographs are also visual, but seem to represent properties that are neither visible nor surface-based. And so on, for a wide variety of visual representations beyond perception. Observations like these make me doubtful about substantive constraints on the types of properties that can be represented. I worry they will inevitably be undercut by human visual ingenuity.

A second and related idea is that for a representation to be visual, its content must exhibit some kind of uniformity in the properties represented across the visual field. For all I have said, a feature map could represent only color in some regions, only edges in others, and only texture in still others. Similarly, it could refer only to points in some regions, surfaces in others, and mid-sized objects in still others. Such contents would seem peculiarly disjointed— perhaps visual in only a partial sense. They appear to violate a norm of visual representation, roughly, that the same type of objects and properties are represented across the visual field. This standard is a reflection of

§7 Further conditions on visual content
the nature of perceptual processing, which characteristically involves parallel computations which involve uniform transformations of the entire visual field. The result is the simultaneous detection of colors everywhere, edges everywhere, and so on.

This standard of uniformity can be thought of as a PROJECTION NORM, because it effectively requires that feature maps be special kind of projection of the centered-worlds where they are accurate. One can imagine a “geometrical-ontological” projection from centered-worlds to feature maps. Such projection would map certain types of objects and properties in the world to feature clusters, according to the directional locations of the map field. The objects and properties mapped would satisfy natural classes like properties of shape, edge, or color, or perhaps classes of higher-level properties like basic kinds (Palmer 1999, ch. 9). Projections like this would provide the type of uniformity sought above. And the variety of possible methods of projection can accommodate the wide variety of types of images, well beyond those involved in perception.

The status of the projection norm is complex. On one hand, projection guarantees the kind of completeness or uniformity which visual representation seems to require. On the other hand, many visual representations appear to be selective and non-uniform in their representation of objects and properties. This is exhibited by many line drawings, where artists selectively render some features in detail, leaving others unspecified or merely sketched. It is also present in mental imagery, which is widely thought to be merely “patchy”, filling in some visual details, but not others. And it is manifested in perception, for example by the object file system, where only a small number of the visually available objects are represented and tracked. (I discuss feature map analyses for both of these phenomena in Section 9.)

A strict enforcement of the projection norm seems too strong as a condition on visual representation; it rules out too many central cases of visual representation. On the other hand, the norm seems to get at an important idea, for it also rules out truly arbitrary distributions of features from the remit of visual representation. I leave the adjudication of this question to future inquiry.

8 Equivalent structures

In this section, I consider two kinds of alternative structure which may carry the same information as a perspectival feature map, one based on purely directional structure, the other propositional. I explain why perspectival feature maps are still the preferred account of visual content, despite the informational equivalence.

According to a PURELY DIRECTIONAL approach, visual contents are identified with a bundle of directions, emanating from an origin point, each of which is associated with a feature cluster.25

---

25According the perspectival feature map analysis, for every object or property represented, that element is necessarily associated with a distinctive two-dimensional segment of the map field, a kind of abstract projected shape. No such structure figures in the purely directional approach. Such segment-shapes in the map field are part of the content itself, on my account; they give a rather literal interpretation to Frege’s idea of a “mode of presentation.”
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 deep 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 loses 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. But such explicit constraints, which would effectively define the location of a map field, make the purely directional approach informationally equivalent to the feature map approach. 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. Here is one rough scheme: for every feature cluster \( \langle a, \{F_1, \ldots F_n\}\rangle \), where \( a \) is an object and \( F_i \) is a property, there is a translation into a conjunctive proposition:

\[
F_1 a \land F_2 a \ldots \land F_n a \ldots
\]

To capture the spatial dimensions of the feature map in propositional form, it is convenient to introduce a coordinate system; then each point \( p \) on the map field can be represented by a set of coordinates, and each direction \( d \) can be represented either as a pair of angles, or as a unit vector \( u_d \). Assuming some translation into a coordinate system, then we may introduce the relation \( R(a, u_d, p) \), that is, that \( a \) is located along the direction of the vector \( u_d \) bound to the initial point \( p \). Now conjoin this proposition to the former: \( R(a, u_d, p) \land F_1 a \land F_2 a \land F_3 a \ldots \). Similar translations may be carried out for the treatment of relations. Then the feature map as a whole may be translated as the conjunction of each such translation of its feature clusters. When a proposition can be translated from a feature map in this manner I will say that the proposition specifies the feature map. The translation scheme offered above is just one among many possible options.

\[26\] Viewpoint-relativity can be handled either in the object language, by introducing lambda abstraction over viewpoints, or in the meta-language, within the definition of accuracy, by assigning a free variable \( v \) to the viewpoint in the index of evaluation.
Given the availability of such propositional translations, one might wonder why attribute feature map structure to visual content in the first place? Why not employ the more familiar propositional structure?

First, as I’ve noted, 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 special reason to expect that their content would exhibit a tree-like structure.

Second, not all propositions specify perspectival feature maps; indeed, only highly specialized propositions do. So even if visual contents were propositional, they would have to be only those propositions which 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 they are ultimately redescriptions of the underlying visual structure.

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 are nevertheless visual. Consider the output of sophisticated computer vision software, like that used in the visual modules of self-driving cars.\(^{27}\) The data structures produced are written in an algebraic code, but nevertheless describe the directions of objects and properties relative to a viewpoint. Despite their obvious differences with pictures and retinotopic representations, such codes nevertheless seem to 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.

To simplify matters, 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. And let us grant that these hold. What then should we say about its visual content? Here, I think, we may postu-
late 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 adopt
a different conception of their content. By describing contents as perspectival feature maps, we
abstract away from possible differences in the underlying substrate, and focus on the representa-
tion’s distinctively visual function.

9 Applications to vision and mental imagery

In this final section I apply the perspectival feature map analysis to a range of special cases in
visual perception and mental imagery. Extensions of the core proposal are introduced along the
way. Though all of the phenomena discussed here are the subject of ongoing empirical study, it is
not my intention to take a stand on open scientific questions. Instead, my immediate philosophical
aim is to show that the visual phenomena in question may, as a matter of principle, be analyzed in
terms perspectival feature maps.

9.1 Feature detection and object files

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 dimen-
sions; 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 ac-
cessible 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 corre-
respond to larger objects. The object file system is thought to 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. Of course, because the science of these representations is itself a work in progress, any philosophical account of their content is conjectural.

To begin, how should we understand the content of a typical Treisman-map? My account requires that every feature cluster contain an object. 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. (Hence it will satisfy the projection norm.) 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. (Hence it will not satisfy the projection norm.) 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 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.

These two examples show the flexibility of the perspectival feature map analysis. For each generates a very different kind of content. Treisman-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. Despite these differences, both representations are visual, and the present analysis helps us understand why.

9.2 Patchy imagery

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 visual-

---

28By contrast, Clark (2004) feature 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.
ize 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 mind’s 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 analysis of image content using the 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 similar 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 resist the projection norm discussed earlier.

9.3 Amodal completion

The perceptual system fills-in a great deal of the world that is not immediately visible on the basis of visual cues. For example, given only visual information from the front-side of a sphere, perception seems to fill in the rest of the volume, attributing to it a spherical shape. In terms of perspectival feature maps, such representations can be understood to contribute rich, three-dimensional shape properties to their feature clusters. A somewhat different kind of case is amodal completion, in which some parts of a contour are visible, some parts are occluded by a second object, and the visual system fills-in the occluded edge (Palmer 1999, ch. 6).

There are two ways one might describe the content of amodal completion in a perspectival feature map. The first is that it works in much the same way as the representation of volumetric solids. A partially occluded sphere is represented by an irregular segment in the map field, but the linked feature cluster contains the property sphere. A different approach holds that, in amodal completion, the completed edges are actually represented by corresponding segments in the visual field, even if they don’t correspond to retinal projections. In this case the partially occluded sphere would be represented by a circular (or ellipsoid) segment on the map field; and it might also
be linked to the property \textit{sphere}. This second approach implies that mereologically independent objects (the sphere and its occluder) can be represented by overlapping segments on the map field. Both construals of the content of amodal completion are in-principle viable, and the best approach will likely be decided by empirical inquiry.

9.4 Shells and volumes

It’s conceivable that the visual system represents both the complete exterior surfaces of objects at some stages, and their fully closed volume at later stages (Marr 1982). Indeed, both could be represented simultaneously, if not perceptually, then in a picture or computer vision. The situation is notable because both the shell of an object and its full shape are associated with the same region of the visual field. How could a perspectival feature map include both without conflating them? The solution is simply to allow that more than one feature cluster may be linked to the same segment of the map field. This was never explicitly ruled out, but we have yet to encounter a use for it. In this case, a single segment would be linked to two feature clusters; one with the shell in the object position and shell-shape in the property position, the other with the volumetric solid in the object position and a closed volume-shape in the property position. The two feature clusters would presumably be linked by some kind of parthood relation as well.

9.5 Transparency

A car is visible through a window; the window is visible through my sunglasses. How should we understand the content of the perception of transparent surfaces? The phenomenon is potentially puzzling in the present framework because it involves multiple perceived objects in overlapping regions of the visual field. The most extreme sort of example, from an experiment by Blaser, Pylyshyn, and Holcombe (2000), seems to elicit the perception of two rotating surfaces, occupying exactly the same position on the visual field, each visible through the other. The solution is to adopt the principle discussed before, that more than one feature cluster can be linked to a single segment. Thus the representation of the window is associated with a segment which is linked to a feature cluster. And within that segment there is a smaller segment which is linked to a feature cluster for the perceived car. Formally, this all proceeds smoothly.

However, it illustrates an important lesson. From the case of shells and volumes, and parts and wholes more generally, one might have expected to find a kind of iconic isomorphism between part-whole relations in the map field itself and part-whole relations among the represented objects. This seems to be the case when representing ordinary part-whole relations, for example. But transparency reveals that this principle is not entirely general. For here, the segment associated with the car is a subregion of the segment associated with the window—yet the car is not part of the window.
9.6 Blur

Blur, whether in vision or photography, seems to involve a kind of imprecision within an image that falls below the level of acuity normally available for that type of image. (In this sense, by contrast, low-resolution digital images are not truly blurry.) Blurry images seem to express content which is correspondingly indeterminate with respect to direction. For example, a crisp representation of an edge locates that edge in a precise direction, while a blurry representation of the same edge seems to locate it in a mere range of directions. In order to capture the effects of blur within the perspectival feature map analysis, one must introduce a level of adjustable imprecision around directions. Though I can imagine various ways of achieving this, here is one. Recall the clause in the definition of accuracy which defines object direction: a segment $S$ is associated with the a feature cluster $C = (O, F^*)$ and a bundle of directions $D$. The definition of accuracy requires that, for every direction $d \in D$, $d$ intersect some part of $O$. A blurry image would require a somewhat different clause, along the lines of: for any direction $d \in D$, there is a $d'$ in an directional interval around $d$, such that $d'$ intersects some part of $O$. Something like this, I believe, would deliver the right kind directionally indeterminate of content. To implement this analysis systematically, some segments would have to be labeled as blurry and associated with a magnitude to indicate the interval of allowed imprecision.

9.7 Time

I have abstracted away from the time dimension throughout this essay. I’ve implicitly assumed that each perspectival feature map represents the world at an instant, and that the centered-worlds relative to which feature maps are evaluated for accuracy are themselves centered at a time. Yet there are a variety of ways that time, or change through time, may be represented visually (McCly 1993, ch. 4). The most basic is through temporal extension of the representation itself, as in the case of perception over an interval, or the continuous run of a camera in film.

The content of such extended visual representations can be understood as densely ordered sequences of feature maps— four-dimensional feature maps, as it were. Each point in such a sequence would correspond to an ordinary, instantaneous feature map, at a time. Together, the entire sequence they would constitute visual content over time. In this setting, it would be natural to allow relations to be linked not only to clusters within a given feature map, but also to clusters in distinct maps. The perception of causation, for example, might express such relations between feature clusters at distinct momentary feature maps.

10 Conclusion

I have outlined a theory of visual content which aims to capture the directional structure of visual space while preserving its basic pictorial organization. Building upon the use of feature
maps in vision science, I’ve proposed that all visual contents are perspectival feature maps, and
developed an account of their structure and accuracy conditions. This proposal, I’ve argued, does
a better job than alternative theories of visual content, and I’ve shown how it may be extended to
a range of real-world cases. By capturing an essential aspect of the structure of visual content, I
hope to have illuminated in part what is distinctive of visual representation.

Yet these conclusions lead directly to new questions: can the contents of other perceptual
modalities, beyond vision, also be understood as feature maps? And if so, what are their dimen-
sions, and what are their core and peripheral features? Further, is there a kind of general feature
map, corresponding to a person’s total perceptual experience or state, which integrates the content
from each modality? I leave these questions to future work, with the hope that the ideas developed
here will usefully generalize to perceptual representation beyond vision.
Appendix: A formal model for perspectival feature maps

I first describe a formal model for perspectival feature maps, then provide a formal definition for accuracy. The definitions offered here are not set within a traditional model theory, but they are intended to be explicit enough so as to facilitate such an analysis as needed.

Definition of perspectival feature maps

A perspectival feature map $M = \langle \text{Space}, \text{Field}, \text{Face}, \text{segments}, \text{Directions}, \text{clusters}, \text{relations} \rangle$, where:

- $\text{Space} = \langle P, d \rangle$ is a 3-dimensional Euclidean space, where $P$ is an uncountable set of points, and $d$ is a distance metric over $P$;
- $\text{Field}$ is the map field, a set of points from $P$ comprising a finite surface within $\text{Space}$;
- $\text{Face} = \langle v_y, v_z \rangle$ is the orientation of the map field, where, if $\text{Field}$ is flat, $v_y$ is a unit vector parallel to $\text{Field}$ and specifies its up-down orientation, and $v_z$ is a unit vector normal to $\text{Field}$ and specifies its front-back orientation;
- $\text{segments}$ is the set of segments in the feature map, where $\text{segments} \subseteq P(\text{Field})$;
- $\text{direction}()$ is a function mapping each point $p$ in $\text{Field}$ to a ray in $\text{Space}$ such that:
  - $\text{direction}(p)$ has $p$ as its initial point;
  - $\text{direction}(p)$ points forward: where $v$ is a vector co-directional with $\text{direction}(p)$, the projection of $v$ onto $v_z$ is co-directional with $v_z$;
  - $\forall S \in \text{segments} : \text{direction}(S) = \{ \text{direction}(p) \mid p \in S \}$;
- $\text{clusters}()$ is a function from segments, elements of $S$, to sets of feature clusters;
  - a feature cluster $C = \langle o, F^* \rangle$, where $o$ is an object and $F^*$ is a set of non-relational and perspectival properties;
  - $\text{object}()$ is a function from a feature cluster $C$ to the object $o$ in $C$;
  - $\text{properties}()$ is a function from $C$ to the set of non-perspectival properties in $C$;
  - $\text{p.properties}()$ is a function from $C$ to the set of perspectival properties in $C$;
- $\text{relations}()$ is a function from sequences of feature clusters $\langle C_1, ..., C_n \rangle$ to sets of relations;
  - $\text{relations}^*(\cdot) \subseteq \text{relations}()$ has only non-perspectival relations in its range; $\text{p.relations}() \subseteq \text{relations}()$ has only perspectival relations in its range.

Definition of accuracy

Next I define accuracy for a perspectival feature map relative to a centered world. I assume a certain amount of geometric and metaphysical structure for possible worlds, including:

- Every object that exists in a world either instantiates or fails to instantiate properties at that
A viewpoint-centered world is a pair \( \langle w, v \rangle \) where \( w \) is a world and \( v \) is a viewpoint within \( w \). A viewpoint \( v = \langle s, L, u_y, u_z \rangle \) (relative to \( w \)), where:

- \( s \) is the projection source; for perspective projections, \( s \in P_w \) is a point; for parallel projections \( s \subset P_w \) is a plane;
- \( L \) is the projection plane; \( L \) is a set of points from \( P_w \) comprising a 2-dimensional plane segment within the space of \( w \);
- \( u_y \) is a unit vector parallel to \( L \) and specifies its up-down orientation; \( u_z \) is a unit vector normal to \( L \) and specifies its front-back orientation.

Next I define the relationship of a congruence between a perspectival feature map \( M \) and a viewpoint \( v \) in \( w \), relative to an isomorphism \( f \) from \( \langle P, d \rangle \), the three-dimensional space of \( M \), to \( \langle P_w, d_w \rangle \), the 3-dimensional space of \( w \). Here I’ll assume that \( \forall s \in \mathcal{P}(P) : f(s) = \{ f(p) \mid p \in s \} \). This isomorphism is understood as a bijection which preserves distance.

A perspectival feature map \( M \) is congruent with \( v \) in \( w \), relative to \( f \) iff:

1. \( f(\text{Field}) = L \);
2. \( f(u_y) \) has the same direction as \( v_y \) and \( f(u_z) \) has the same direction as \( v_z \);
3. \( \forall p \in \text{Field} : f(\text{direction}(p)) \) is collinear with the line that intersects \( s \) and \( f(p) \).

We are finally in a position to define accuracy for perspectival feature maps. Here I assume a viewpoint appropriate for perspective projection.

A perspectival feature map \( M \) is accurate at a viewpoint-centered world \( \langle w, v \rangle \) iff there is an isomorphism \( f \) from \( \langle P, d \rangle \) to \( \langle P_w, d_w \rangle \) such that:

(a) \( M \) is congruent with \( v \) relative to \( f \);
(b) for all \( \langle C_1, ..., C_n \rangle \) such that \( \forall i : 1 \leq i \leq n : \exists S \in \text{segments} : C_i \in \text{clusters}(S) \) and
   1. \( \forall r \in \text{directions}(S) : \text{some part of object}(C_i) \) intersects \( f(r) \) in \( w \);
   2. \( \forall F \in \text{properties}(C_i) : \text{object}(C_i) \) is \( F \) at \( w \);
   3. \( \forall G \in \text{p.properties}(C_i) : \text{object}(C_i) \) is \( G \) relative to \( v \) at \( w \);
   4. if \( \langle C_1, ..., C_n \rangle \in \text{domain}(\text{relations}(\)) \):
      \( \forall R \in \text{relations}^* \langle C_1, ..., C_n \rangle : \text{object}(C_1), ..., \text{object}(C_n) \) stand in \( R \) at \( w \);
      \( \forall Q \in \text{p.relations}(\langle C_1, ..., C_n \rangle) : \text{object}(C_1), ..., \text{object}(C_n) \) stand in \( Q \) to \( v \) at \( w \).
References


References


Treisman, Anne (1986). “Features and objects in visual processing”. In: *Scientific American* 255.5, 114B–125.


