Contours of Vision: Towards a Compositional Semantics of Perception

Kevin J. Lande

ABSTRACT

Mental capacities for perceiving, remembering, thinking, and planning involve the processing of structured mental representations. A compositional semantics of such representations would explain how the content of any given representation is determined by the contents of its constituents and their mode of combination. While many have argued that semantic theories of mental representations would have broad value for understanding the mind, there have been few attempts to develop such theories in a systematic and empirically constrained way. This paper contributes to that end by developing a semantics for a 'fragment' of our mental representational system: the visual system's representations of the bounding contours of objects. At least three distinct kinds of composition are involved in such representations: 'concatenation', 'feature composition', and 'contour composition'. I sketch the constraints on and semantics of each of these. This account has three principal payoffs. First, it models a working framework for compositionally ascribing structure and content to perceptual representations, while highlighting core kinds of evidence that bear on such ascriptions. Second, it shows how a compositional semantics of perception can be compatible with holistic, or Gestalt, phenomena, which are often taken to show that the whole percept is 'other than the sum of its parts'. Finally, the account illuminates the format of a key type of perceptual representation, bringing out the ways in which contour representations exhibit domain-specific form of the sort that is typical of structured icons such as diagrams and maps, in contrast to typical discursive representations of logic and language.

1 Semantics for Mental Representations

A foundational principle of cognitive science is that mental capacities for perceiving, remembering, thinking, and planning involve the formation and transformation of representational mental states. To be in my current visual state is in part to represent the apple in front of me as round and red. This visual state has semantic properties: it is about the apple, and it is accurate to the degree that the apple actually is round and red. Another foundational principle is that these mental representations are 'selectively organized data structures', as the psychologist Stephen Palmer ([1977], p. 442) put it. Plausibly, my visual representation of the apple has as constituents a representation of the apple's shade of red and a distinct representation of its round shape. My visual representation therefore has what we may call 'structural' or 'syntactic' properties corresponding to its constituent representations and their modes of combination (Marr, [1982], pp. 20–1).

Tying these together, a third principle is that mental representations are 'compositional': the way a complex representation represents the world is exhaustively a function of the way that its constituents represent the world and of how those constituents are structurally related (Fodor and Pylyshyn, [1988]). My visual state represents the apple as red and round because it is made up of a constituent representation of that shade of red and a constituent representation of that round shape, and because these constituents are combined with each other in a certain way. Capacities to represent an infinitely rich and varied range of contents emerge, all as a package, from primitive capacities to represent basic contents and from capacities to combine these representations. These principles provide a framework for explaining how exceptionally rare configurations can be coded effectively in terms of exceedingly commonplace elements.

If these principles are correct, then it should be possible to formulate systematic semantic theories of mental representational systems, much as linguists formulate semantic theories for fragments of natural languages. A semantic theory is a theory of representation: it specifies what representations there are in a system and what they function to represent. A compositional semantics in particular aims to identify (a) the system's primitive representations and their contents, (b) the structural principles by which representations in the system enter into combinations with each other, and (c) the semantic import of those structural principles—i.e. how the contents of the constituents determine the contents of the whole, given the way those constituents are combined. Such a theory purports to explain a system's capacity to represent a range of contents, abstracting from the circumstances under which the system forms those representations.

A semantics for mental representations would have significance for understanding the computational architecture of the mind, the epistemic relations between mental states, and the grounds on which mental states have the contents that they do. Philosophers of mind have long been concerned with foundational questions about the general shapes that

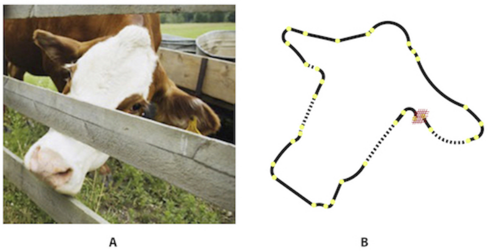


Figure 1: Theories of contour perception hold that the visual system codes the outline shapes of objects such as the cow's head in (A) by a combination of representations of distinct contour segments, as illustrated in (B). From (Kellman, Garrigan and Erlikhman, [2013], p. 271) [reprinted with permission].

semantic theories of perception and thought should take. But few attempts have been made to build on those foundations so as to construct systematic, empirically constrained semantic theories of these representational systems.

Since semantic theories aim to characterize real representational capacities, there should be 'factual considerations which constrain theories about the internal code', as Jerry Fodor ([1975], p. 99) suggested. Psychology is replete with theories of representational processes. But as in philosophy, few theories undertake a systematic analysis of the representations that enter into those processes—their structure and content. The vision scientist Jacob Feldman ([1999], p. 211) noted that '[v]ision researchers have hardly ever touched on semantics in an explicit way', while Barbara Von Eckardt ([2012], p. 33) writes, '[t]here is nothing even approximating a systematic semantics for even a fragment of [the mental representational system]'. In psychology, representational theories of process are common; theories of representation are scarce.

This paper aims to contribute to the semantics of mental representations, conceived of as a constructive empirical enterprise. As a case study, I focus on representations in 'mid-level vision' of the bounding contours of objects, such as the outline of the cow's head pictured in Figure 1a. Perceiving an object's boundary can be critical to segregating that object from its background and to representing its volumetric shape (Koenderink, [1984]) and its category (Biederman and Ju, [1988]) at a given moment. Psychological models standardly take contours to be represented in a structured code: representations of outline contours are organized from representations of distinct contour segments, such as the segments illustrated in Figure 1b, which in turn are organized from representations of various features such as orientation, size, and curvature.

Object boundaries can be irregular; they can overlap, intersect, and interrupt each other. Their fragments don't come to the visual system stamped with the identities of the others with which they belong. Vision must determine how the fragments can and cannot go together, so as to compose a representation of an integrated boundary. The psychologist Brian Keane ([2018], p. 5) writes, 'Just as linguists must work hard to discern the eligible phonemes and computational rules that ultimately lead to the well-formed syntax of a native speaker, so too must vision scientists carefully design experiments to figure out the features and compositional rules that govern [contour representations]'.

Sections 1–3 develop an account of how representations of scenes containing object

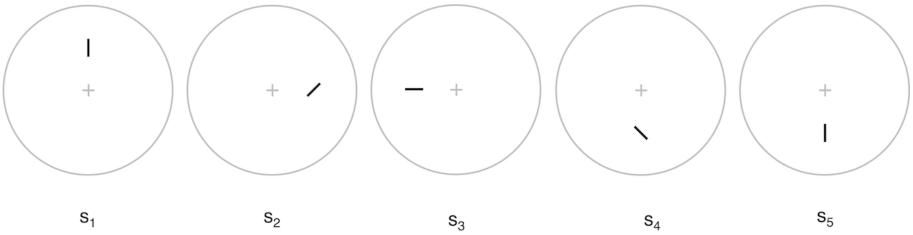


Figure 2: Isolated segments of object boundaries. The central cross indicates the point at which the viewer's gaze is fixated and the surrounding circle indicates the subject's field of view; neither are part of the stimuli.

contours are composed from representations of more elementary contour segments and their features. I survey evidence that the perception of the bounding contours of objects in a scene involves at least three syntactically and semantically distinct modes of composition: 'concatenation' (Section 2), 'feature composition' (Section 3), and 'contour composition' (Section 4). The account developed in these sections serves more generally to model a working framework for formulating and evaluating compositional analyses of perceptual representations. In Section 5, I consider a broad challenge for analyses of this sort. Holistic, or Gestalt, perceptual effects are pervasive throughout perception, including the perception of object contours. These effects are widely taken to demonstrate that the whole percept is 'other than the sum of its parts'. I show how the account developed here can handle such effects, demonstrating that the holistic character of perception is compatible with its compositionality. Finally, in Section 6, I consider implications of the present account for the format of contour representations. Contour representations exhibit what I call 'domain-specific form,' which is a hallmark of structured icons such as diagrams and maps and contrasts with typical representations of logic and language.[1]

2 Arrays of Elements: Perceptual Concatenation

I now introduce a working framework for formulating and evaluating hypotheses about the compositional structure and content of perceptual representations. As a starting point, I focus on representations of fragments of object contours and of arrays of such fragments. The next two sections introduce important modifications to this account.

Consider representations of highly simplified stimuli in which an isolated segment of an object's contour is visible, as if the rest of the object has been masked (as depicted in Figure 2).[2] These elements can have one of four positions (top, bottom, left, right) and one of four orientations (vertical, horizontal, tilted right, tilted left). Distal object contours are not to be confused with the proximal patterns of light that those contours project to the retina, which we might call their 'retinal silhouette'. Following Casati and Varzi ([1999], Chapter 5), I assume that object contours and their segments are spatially located entities.

---

[1] The immediate target of this discussion is the structure and content of potentially non-conscious representations in visual information-processing, rather than the structure and content of conscious visual phenomenology. However, I take it that the latter depends at least in part on the former.

[2] For now, I assume that representations of these isolated elements function to single out parts of an object's bounding contour. I motivate this assumption in Section 4.

Contours are parts of the objects that they bound. They can change positions both when the whole object moves and when it is warped. Contours instantiate various features, such as shape, orientation, curvature, size, and contrast (Davies, [2020]). These features can change over time. The existence and features of an object's boundary are independent of perceiver and viewpoint. As a coin rotates in depth, it maintains its circular boundary while its retinal silhouette takes various elliptical shapes.

Plausibly, semantic concepts such as 'accuracy' and 'reference' play an explanatory role in characterizing representations of distal stimuli, but not in characterizing mere sensory registrations of proximal stimuli (Burge, [2010]). Accordingly, despite the simplicity of the examples, my concern throughout will be with perceptual representations of distal object contours rather than low-level sensory responses to retinal silhouettes.[3]

I use boldface letters to name perceptual representations. Let $\mathbf{s_1}, \mathbf{s_2}, \mathbf{s_3}, \mathbf{s_4}$, and $\mathbf{s_5}$ denote one's perceptual representations of the distal stimuli in Figure 2—s_1, s_2, s_3, s_4, and s_5, respectively. I will characterize these representations as having accuracy conditions of the following sort:

(s_1) If $\mathbf{s_1}$ is a representation of x in context δ, then it is perfectly accurate if and only if x at least partially bounds an object and is located at top of scene and frontally oriented 0° with respect to the scene's vertical axis in δ.

More generally, for a representation $\boldsymbol{\alpha}$,

(A) If $\boldsymbol{\alpha}$ is a representation of x in δ, then it is perfectly accurate iff $\ldots x \ldots$.

Semantic clauses of this kind do not describe a visual system's input-output mapping from stimuli to representations; rather they specify the contents of possible output representations. The right-hand side of the clause says how the representation characterizes things, by specifying the conditions ('$\ldots x \ldots$') that must be met for the representation to be veridical, where the relevant species of veridicality is accuracy. The accuracy of a representation depends on whether the representation's target or referent ('x') in a given context ('δ') satisfies the relevant conditions. I abstract from details of what is referred to in any particular context by considering conditionalized accuracy conditions, which specify what the accuracy conditions of a representation would be were it to refer to a given particular in a given context.

A perceptual representation can be designated (substituting in for '$\boldsymbol{\alpha}$') with either a primitive name (for example, '$\mathbf{s_1}$') or else, as we will see, a structural analysis of the representation. A representation's accuracy conditions do not determine its structure. (A) contains placeholders for a representation's referent, context, and the condition it places on its referent, but it does not follow that the representation itself has distinct constituents dedicated to representing the referent x, context δ, and condition that x must fulfill in δ.

(A)-style clauses provide a tentative, working framework for semantically analyzing perceptual representations. I now turn to highlighting core empirical, explanatory roles of different elements of these analyses.

---

[3] Likewise, while stimuli such as those in Figure 2 have been used to characterize the receptive fields of simple 'retinal edge detector' cells in early visual cortex (Hubel and Wiesel, [1968]), I am concerned here with representations that emerge from neuronal activity that exhibits selectivity for object contours with specific environmental orientations and positions in depth (Sauvan, [1999]).

2.1 Accuracy Conditions

I take perceptual representations to be causally efficacious states (or episodes) of an individual or an individual's perceptual system, which function to—well—represent. Instead of giving a definition or constitutive account of 'representation', I will discuss how positing representations with specific accuracy conditions helps to explain certain patterns of psychological evidence. I focus on the role of representations in 'perceptual constancies'—roughly, capacities to represent distal features of items despite significant variability in proximal stimuli (Burge, [2010]).

Suppose that an observer is repeatedly presented with a line segment of varying orientation and is tasked with selecting from a set of comparison line segments the one that most closely matches the target. And suppose that the subject's head is tilted at different angles across trials, or that the line segment's three-dimensional orientation in depth varies as well as its two-dimensional orientation on the retina. Under a broad range of circumstances, it is the orientation of the target with respect to the environment that strongly predicts the orientation of the selected comparison, while there may be no correlation between their retinal orientations (Wade and Swanston, [2013], pp. 181–87, 255–59). Normally, a lamppost will appear to be standing upright no matter how one's head is tilted. This pattern of results replicates across tasks, types of responses, and under a variety of viewing conditions.

To be sure, one could in principle predict the orientation of the selected comparison as some function of proximal features (such as the conjunction of retinal orientations and the activity of otolith receptors in the vestibular system) that are 'cues' to the distal item's orientation under certain conditions. Normally, however, different cues are operative in different conditions, and of the infinitely many possible mathematical functions that can be defined over these proximal features, the functions that are predictive of behaviour are just those the values of which roughly correspond to the distal item's environmental orientation across the relevant range of conditions. Normally, distal orientation is a better predictor of task responses across conditions than any independently individuated function of proximal features.

Even if the distal target's environmental orientation is predictive of responses for a variety of tasks and conditions, that item's orientation is not sufficient to causally explain the subject's responses. Any effect that the distal object has on the subject's responses must be mediated by proximal sensory receptors, such as photoreceptors and otolith receptors. Indeed, in the best conditions the distal orientations of the targets and the selected comparisons are only approximately correlated. At worst, there are conditions in which the correlation breaks down entirely.

How can a distal item's orientation be so predictive of responses under a range of conditions if it cannot causally explain those responses? A natural explanation of this relationship between the target's orientation and the perceiver's behaviour is that this behavior causally depends on internal psychological 'representations', or estimates, of the target's environmental orientation. That these internal representations are presumed to be latent causal variables in the perceiver's psychology explains why different types of tasks and responses conform to a similar pattern: these task responses all causally depend on a common type of internal state. That these states are representations of the target's environmental orientation explains why this is a pattern in which the orientation of a distal item predicts responses better than any independently individuated function over proximal features. Other things equal, the degree to which the state accurately represents the orientation of the distal item determines the degree to which the distal item's actual orientation predicts task responses.

Here, then, is a standard pattern of inference in perceptual psychology: (i) It is observed that, under certain conditions, features of the distal stimulus predict task responses better than any independently individuated feature of the proximal stimulus; (ii) Task responses cannot causally depend on the distal stimulus itself; (iii) Part of an explanation of the observed pattern is that task responses causally depend on an internal representation or estimate of the relevant features of a distal item. Other things equal, that item's features predict task responses because, and to the degree that, the representation is about that item and accurate about those features.

This pattern of explanation provides a core (but not exhaustive) ground for positing representations with specific accuracy conditions, and has led perceptual psychologists to posit visual representations of intrinsic size, shape, location in depth, and so on. Such representations are themselves targets of causal explanation, for we want an account of how they are produced from stimuli and how they influence other mental states. But our focus here is on the role that positing representations can play in explaining the relations between distal conditions and a subject's responses.

The notion of accuracy occupies a central place in this type of explanation. Other things equal, the more accurate the representation, the more the distal feature of the target will predict the subject's response on the task in relevant conditions. Accuracy comes in degrees. Models of perception regularly depend on psychometric functions that specify not just whether a given type of perceptual state is likely to be accurate in a given condition, but how accurate it is likely to be. In specifying conditions for perfect accuracy, semantic clauses of the form in (A) provide an anchor point for more general measures of degrees of accuracy.[4]

Reference also plays a role in this type of explanation. To evaluate the accuracy of a perceptual representation one normally must, at least implicitly, identify what entity (if anything) it is a representation of. The particular target or referent of the representation, which supplies the value of 'x' in the accuracy condition, constitutes the standard against which the accuracy of a perceptual estimate is measured. The accuracy of one's representation of s_1, for example, depends on the difference between the represented orientation and position and the actual orientation and position of the target element in s_1 that one is viewing. For a particular item to be the referent with respect to which the accuracy of a representation is evaluated, it is not necessary that the representation uniquely characterize that item or that the perceiver attend to or notice it. What a representation is of can depend on contextual factors outside of the perceiver's psychology, such as the causal role that a distal object played in the formation of the representation.

I will accommodate the context-sensitivity of perceptual reference by specifying 'conditionalized' accuracy conditions, which determine the accuracy conditions of a representation as a function of which item the representation is about in a given context (Larson and Segal, [1995], Chapter 6). I set aside questions about what the referent of a representation is in a given context and how it is fixed (see Lande, [2021b]).

---

[4] The discussion here is neutral as to whether perceptual representations also have propositional truth conditions in addition to accuracy conditions.

2.2 Structural Analysis

A structural analysis of a representation specifies how that representation is organized from primitive constituents that stand in certain relations to each other. To specify a representation's structure, it is not enough to specify its accuracy conditions. A representation's structure is a further empirical fact. For example, suppose the stimulus domain includes arrays of multiple distinct line segments. Even if a representation of such an array refers to multiple separate items, and even if it is accurate only if there are multiple such items in the scene, it does not follow that those items receive distinct constituents in the representation. According to some models, texture perception represents arrays of texture elements in terms of their summary statistics (for example, average orientation and distribution), without forming separate representations of the individual elements (Portilla and Simoncelli, [2000]). What sort of evidence would show that different elements in an array receive distinct representations?

I have argued (Lande, [2021a]) that structural analyses of mental representations function to explain the 'distribution' of those representations in a system—namely, what types of representations are psychologically possible in a given system and how these representations can, cannot, or must co-occur. Controlled psychophysical experiments can provide evidence that the visual system is capable of forming some representations but not others, or that some representations can only be formed alongside certain others. The structural analysis of a perceptual representation specifies how that representation is structurally possible in a given system and how it can pattern with other representations in that system. Holding certain other factors equal, the structural (im)possibility of a representation explains the psychological (im)possibility of that representation.

Patterns of statistical co-variation among perceptual responses provide a key source of distributional evidence. Consider the psychologist Wendall Garner's (1974) speeded classification paradigm. Suppose a subject is presented with stimuli containing one oriented contour fragment on top and another on bottom, with the task of detecting as quickly and accurately as possible whether the top item is vertical or not. In a 'baseline' condition, the target feature (the orientation of the top element) randomly varies across trials, while the non-target feature (the orientation of the bottom element) remains fixed. In a 'filtering' condition, the non-target feature varies as well.

If the visual system simply gives a primitive label to each whole configuration, then responses in the filtering condition, which varies between sixteen configurations, should be more variable—slower and less likely to be accurate—than responses in the baseline condition, with varies among only four configurations. Suppose, by contrast, that performance in the filtering condition does not differ significantly from baseline performance. In other words, the increased variability in the bottom element has little effect on discrimination of the top element—responses to the top element are statistically independent of responses to the bottom one. Part of the explanation for this independence in responses would be that those responses depend on separate representations. If the top item is represented by a different variable in the visual system than the bottom item, then the visual system can marginalize over one variable, insulating responses to the one element from variability of other parts of the scene (Algom and Fitousi, [2016]).

Indeed, in some conditions, as when the elements in the array are small and densely packed, responses to arrays of elements may be highly co-variant—consistent with the elements' being represented together as a texture—while in other conditions the visual system treats some elements in an array relatively independently (Kimchi, [1992]).

Let's focus on cases in which different line elements in an array receive distinct representations, such that responses to those elements can be relatively independent of each other. One can model the representation of a two-element display as a set of distinct representations, $\{\mathbf{e_a}, \mathbf{e_b}\}$, where $\mathbf{e_a}$ and $\mathbf{e_b}$ each represent a different element as having a certain position and orientation. That the representation, $\{\mathbf{e_a}, \mathbf{e_b}\}$, of the whole stimulus has as constituents representations, $\mathbf{e_a}$ and $\mathbf{e_b}$, of the two elements captures the fact that the representation of the whole stimulus requires representing the orientations of each element. That each element receives a distinct constituent captures the fact that responses to one element are independent of responses to the other.

More generally, let REP be the set of all structurally possible visual representations as of contours that either fully or partially bound objects. A basic goal of a syntactic theory of such contour representations would be to specify which representations belong in the set, REP, and how they can be distributed. A syntax for the set would identify some primitive representational elements in the set. Every other representation in the set can be individuated as a function of a sequence of structural operations, or modes of combination, applied to those primitives. These structural operations need not correspond to causal and temporal relations between representations. An account of the way a given representation is structurally related to its constituents explains why that representation is an eligible member of the set of possible representations in the system, not how or under what conditions the representation is formed.

Let '+' be a mode of combination, or structural operator, that maps representations $\alpha_1, \ldots, \alpha_n \in \text{REP}$, to a complex representation, $\ulcorner\{\alpha_1, \ldots, \alpha_n\}\urcorner \in \text{REP}$.[5] Call this mode of combination, 'perceptual concatenation'. Concatenation forms a representation with the structure of a set. For present purposes, suppose that other things (limitations of attention, memory, and noise) equal, one can visually represent arbitrary arrays of elements in a non-texture-like way, such that the representations of each element can vary more or less independently of the others.[6] Then concatenation can be defined as follows:

> Concatenation (SYN): If $\alpha_1, \ldots, \alpha_n \in \text{REP}$, then $\ulcorner\alpha_1 + \ldots + \alpha_n\urcorner \in \text{REP}$, where $\ulcorner\alpha_1 + \ldots + \alpha_n\urcorner = \ulcorner\{\alpha_1, \ldots, \alpha_n\}\urcorner$.

This says that any concatenation of multiple representations is itself a representation, and that this representation has the structure of an unordered set.

2.3 Compositional Semantics

I have indicated what kinds of psychophysical evidence provide reasons to posit representations, for example of contour fragments, with specific accuracy conditions (Section 2.1) and what kinds of evidence motivate specific structural analyses of representations,

---

[5] Here 'α_i' are schematic variables, for which one can substitute either structural descriptions or names of perceptual representations, such as '$\mathbf{e_a}$' or '$\mathbf{e_b}$'. I use corner quotes, or quasi-quotes, around complex expressions that contain free schematic variables, for example, $\ulcorner\alpha_1 + \alpha_2\urcorner$. These quasi-quoted expressions designate the form of a representation while abstracting from the identity of its constituents.

[6] There may be capacity limits on how many contours can in fact be represented independently at a given time, but such processing limits are not typically treated as constraints on which representations are structurally possible (Lande, [2021a], pp. 662–63).

for example of arrays of contour fragments (Section 2.2). But I have not said anything about the accuracy conditions of these latter complex representations. I now want to show how one can combine semantic and structural hypotheses in order to project the accuracy conditions of arbitrary complex representations.

To begin with, we can recast (s_1)–(s_5) as giving the accuracy conditions for atomic representations of contour elements:

Atoms: (1) $\{\mathbf{e}_{\mathbf{t0}}, \mathbf{e}_{\mathbf{t45}}, \ldots, \mathbf{e}_{\mathbf{r0}}, \ldots, \mathbf{e}_{\mathbf{l135}}\} \subseteq \textsc{rep}$.

(2) If $\mathbf{e}_{\mathbf{t0}}$ is a representation of x in δ, then it is perfectly accurate iff x at least partially bounds an object and is located at top of scene and frontally oriented 0° w.r.t. the scene's vertical axis in δ.

⋮

(17) If $\mathbf{e}_{\mathbf{l135}}$ is a representation of x in δ, then it is perfectly accurate iff x at least partially bounds an object and is located at left of scene and oriented 135° w.r.t. the scene's vertical axis in δ.

A concatenation of perceptual representations is accurate to the degree that all the concatenated constituents are accurate:

Concatenation (SEM): $\ulcorner \alpha_1 + \ldots + \alpha_n \urcorner \in \textsc{rep}$ is perfectly accurate iff for every constituent, $\alpha_{i:1\leq i\leq n}$, α_i is perfectly accurate.

For example, if the task is to match one array of oriented elements to another, then, other things equal, success on the task depends just on how well one independently discriminates the orientation of each element.

We now have a rudimentary compositional semantics for representations of arrays of contour elements. (1)–(17) constitute the atomic base of the theory, while the syntax and semantics of concatenation constitute the compositional assembly. The semantic clauses, (2)–(17) and Concatenation (SEM), should figure into the best explanation of why under certain conditions the environmental orientation and position of a line element (or array of such elements) predicts perceptual responses better than any independently specifiable features of the proximal stimulus. Structural analyses of representations in terms of (1) and Concatenation (SYN) should figure into the best explanation of why the perceptual effects of varying one edge element are relatively independent of the perceptual effects of varying the other. The pair of Concatenation (SYN) and Concatenation (SEM) explain, respectively, how a given concatenated representation is structurally possible and why it has certain accuracy conditions.

3 From Features to Fragments: Feature Composition

While concatenation may help explain our abilities to represent an array of multiple contour fragments in a given scene, it does not account for the way we code complex, extended contours in terms of simpler fragments, or for how those simpler fragments are coded in terms of different features. Let's begin with the latter issue.

The received view in vision science is that representations of even simple forms are not primitive, but consist of representations of the position, orientation, curvature, size, contrast, and so on, of a given element. Such a hypothesis best explains the patterns of variability in how we perceive contours. I will outline the constraints on and semantic import of a distinct mode of combination, 'feature composition', which relates the representation of a contour element to the representations of that element's features. I will argue that feature composition should not be reduced to concatenation.

3.1 Features of Fragments

Capacities to represent the orientation, position, curvature, contrast, and other features of contours are relatively independent. This claim is not based on the fact that contours have these different features, or even on the fact that contour representations have content about these different features. Rather, that the feature representations themselves are distinct—that the features are coded by distinct 'variables' in the visual system—can be seen by looking at patterns of independent variability in how perceivers respond to such features.

For example, the precision with which one represents a line segment's orientation can vary independently of the precision with which one represents its position. If one's task is to discriminate the position of a line segment, then randomly varying the orientation of the element will have marginal influence on one's performance, and *vice versa*. Part of the explanation of this statistical independence is that 'humans encode position and orientation information in markedly distinct manner' (Christensen et al., [2019], p. 18). Likewise for other features such as size, curvature, and contrast polarity (whether the contour is on a light surface against a dark background or *vice versa*).[7]

'Illusory conjunctions' further demonstrate the distinctness of representations of position, orientation, and so on. An illusory conjunction occurs when one accurately represents the features and positions of different elements in a display, but misrepresents which features and positions are co-instantiated in a particular item (Treisman and Schmidt, [1982]). For example, if one views a black vertical element on top and a white tilted line on bottom against a medium gray background, one might misperceive one item as a white vertical line on top and the other item as a black tilted line on bottom (swapping contrast polarities across items), or one might misperceive one item as a black tilted line on top and the other item as a white vertical line on bottom (orientation swapping). Crucially, these errors characteristically involve accurately representing features and locations that items in the scene actually have, but attributing those features to the wrong items.

A standard explanation of this pattern of errors is that the features and positions of the different items receive distinct representations and that this same set of representations can be combined in a number of different ways (Matthen, [2005]). Illusory conjunctions are inaccurate representations of elements in the scene that can nevertheless have more or less accurate feature representations as constituents.

That a contour's position, orientation, colour, and size—say—receive distinct representations does not imply that any of these feature representations can occur in isolation.

---

[7] The hypothesis that an item's different features receive distinct representations is compatible with different neurophysiological possibilities. One possibility is that there be distinct neural populations dedicated to encoding those distinct features. Alternatively, information about position and information about form can be carried in different aspects of a single cell's response (Pasupathy et al., [2018]). Or content about different features may be decodable from separable patterns of activity across ensembles of cells (Christensen et al., [2019]; Taylor and Xu, [2022]).

In early work, Anne Treisman conjectured that 'features of unattended objects may be free floating spatially [...] Locating a feature would, on this hypothesis, be a separate operation from identifying it' (Treisman and Gelade, [1980], p. 100). On one interpretation of the 'free-floating feature' hypothesis, the combination of orientation and position representations is optional and depends on the deployment of spatial attention.

Subsequent studies have shown that if subjects are able to represent an item's orientation (or colour, or other such features), then they are also able to represent its location to varying degrees of precision (Quinlan, [2003]). Later accounts of feature integration incorporated the view that feature representations are attached mandatorily to position representations, independently of attention. The 'feature map' hypothesis preserves the hypothesis that an item's position, orientation, colour, size, and so on are represented independently by different variables in the visual system, but holds that these representations, and any combination thereof, must occur alongside an assignment of position to the item that has the given feature(s) (Clarke, [2021]).

Let's distinguish 'atomic' representations, which are the least complex representation that can occur in isolation, from their sub-atomic primitives, which which can vary more or less independently of each other but cannot occur in isolation. Let $\textsc{contour} \subseteq \textsc{rep}$ be the set of contour representations. Let $\textsc{pos} \subseteq \textsc{rep}$ be the set of sub-atomic position representations, $\textsc{orient} \subseteq \textsc{rep}$ the set of sub-atomic orientation representations (and so on for size, colour, curvature, etc.). Finally, let 'feature composition' refer to the way atomic representations of contour elements are composed from sub-atomic representations of those elements' positions and features.[8]

Let's say that atomic contour representations have the form of a vector, $\ulcorner\langle \alpha_1, \ldots, \alpha_n \rangle\urcorner$, the components of which correspond to different sub-atomic representations of an item's features. A constraint on the construction of such vectors is that they must always contain a component dedicated to representing position:

> Feature Composition (SYN): $\ulcorner \alpha_1 \times \ldots \times \alpha_n \urcorner \in \textsc{contour}$, where $\ulcorner \alpha_1 \times \ldots \times \alpha_n \urcorner = \ulcorner\langle \alpha_1, \ldots, \alpha_n \rangle\urcorner$, iff $\alpha_1 \in \textsc{pos}$, and each $\alpha_{1<i\leq n}$ is from a distinct type of feature representation (drawn from either $\textsc{orient}$, or $\textsc{size}$, or etc.).

This says that feature composition produces well-formed representations of contour segments, that it must always involve a combination of feature representations with a positional representation, and that multiple distinct features of the same type (for example, orientation) cannot be integrated concurrently.[9]

These constraints imply that feature composition is not a mere concatenation of representations into a 'bag of features' representation. There are substantive constraints on how feature representations can combine.

3.2 The Semantics of Feature Composition

Unlike concatenation, feature composition has the semantic import not just that the constituent feature representations be accurate of their respective targets, but that they be

---

[8] I distinguish 'feature composition', which refers to a structural relationship between a representation and its constituents, from 'feature integration' or 'feature binding', which are commonly used to refer to a causal process that generates such representations.

[9] Feature composition applies not just to representations of contours but also representations of the surfaces, parts, or objects that those contours bound. These are outside the scope of the current account.

accurate of the same target. One way for an atomic representation to be inaccurate is for it to misrepresent either the position or orientation of the item that it represents—that is, for one of its constituent representations to be inaccurate. But another way for the atomic representation to be inaccurate is for it to be a combination of an accurate representation of how one item is oriented and an accurate representation of how a different item is located (or sized, or coloured, etc.). This is what can happen in illusory conjunctions.

There are multiple ways to account for the semantic import of feature composition. One route would be to posit additional complexity in the representations. For example, suppose that the sub-atomic representation of an item as having a vertical orientation has the structure $\mathbf{0}(\mathbf{x})$—with a referring representation, $\mathbf{x}$, as one constituent, and an attributive representation, $\mathbf{0}()$, which is accurate just of vertical elements, as a separate constituent. Perhaps the mode of composition is something like function application (Heim and Kratzer, [1998]). Likewise, perhaps the sub-atomic representation of an item as being located in the top position has the structure $\mathbf{t}(\mathbf{y})$, where $\mathbf{y}$ is a referring variable and $\mathbf{t}()$ is an attributive constituent that is accurate just of elements in the top position.

Supposing that there is some way of representing identity in perception, one might propose a reduction of feature composition to the concatenation of different feature representations with an identity representation. That is, to describe a representation as having the structure $\mathbf{0}(\mathbf{x}) \times \mathbf{t}(\mathbf{y})$, is really to ascribe to it the following structure: $\mathbf{0}(\mathbf{x}) + \mathbf{t}(\mathbf{y}) + \mathbf{x} = \mathbf{y}$. Alternatively, one could take feature composition to be analogous to joint predication, which takes multiple predicate-like constituents, $\mathbf{0}()$ and $\mathbf{t}()$, and jointly applies them to the same referring representation, $\mathbf{x}$ (Clark, [2004]).[10]

It may well be that operations such as function application, identity, or joint predication must figure into the specification of the accuracy conditions of certain representations. But such semantic considerations do not determine the structural analyses of the representations themselves. One needs independent evidence that the representation of an element as vertical decomposes into an attributive representation and a separate variable-like referring expression. As Ned Block ([2023], p. 217) points out, the referential aspect of a representation's accuracy conditions may be grounded in aspects of the representation's systematic functional role without the representation's having a distinct constituent dedicated to bare referring.

The simplest interpretation of the evidence discussed here is that feature composition is a primitive mode of combination, not to be analysed in terms of concatenation or function application, and that it has the following semantic import:

> Feature Composition (SEM): If $\ulcorner \alpha_1 \times \ldots \times \alpha_n \urcorner \in$ CONTOUR is a representation of x in δ, then it is perfectly accurate iff
>
> (a) x at least partially bounds an object in δ, and
>
> (b) every constituent, $\alpha_{i:1\leq i\leq n}$, is perfectly accurate in δ, and
>
> (c) for all y, if there is a constituent, $\alpha_{i:1\leq i\leq n}$, that is a representation of y in δ, then $x = y$.

The accuracy of the complex representation requires that the constituents be about the same element. This is a commitment that the constituents themselves do not have.

---

[10] Note that these proposals, as stated, fail to capture the syntactic constraints on feature composition—for example, that position representations are mandatory constituents of contour representations.

Feature composition therefore introduces 'new' content, we might say, which goes beyond the contents of its constituents, in the sense that the accuracy of the constituents is not sufficient for the accuracy of the whole. This is a further way in which feature composition is semantically unlike concatenation.

At the same time, there is no need to posit variable-like representations, representations of identity, or an analogue of joint predication in order to secure this commitment of identity. That commitment is baked into the mode of composition. All this requires is that the visual system have a distinctive way of packaging together feature representations and that the function of this packaging is to represent an item as jointly having all the represented features.

3.3 Analog Primitives

While my focus is on the compositional aspect of perceptual representations, I'll pause to say something about the primitive representations themselves. I assume that these representations are analog in the sense that primitive representations stand in psychological relations that function to mirror objective relations among the features represented, and that this correspondence helps to determine the accuracy conditions of those primitives (Gauker, [2012]; Beck, [2019]).[11] The psychological similarity between representations, for example, can be operationalized by how easy it is to confuse one for the other. Just as a 90° orientation is more dissimilar to 0° than 45°, the representation **90** as of a contour oriented 90°, is more psychologically dissimilar to the representation **0** as of a contour oriented 0° than to the representation **45** as of a contour oriented 45°.

Let μ be a homomorphism from primitive feature representations to distal features of the world, such that there is a correspondence between relevant psychological relations $\mathcal{R}$ among representations and relevant objective relations $\mathcal{R}'$ among represented distal features. Instead of separately specifying the accuracy conditions of each orientation representation, one can just specify that an orientation representation, θ, is perfectly accurate just in case the item that it represents has an actual orientation $\mu(\theta)$, and likewise for primitive representations of position. In this way, we can replace the clauses for primitive feature representations, (1)–(17), with clauses for sets of representations that fall under a relevant psychological relation:

Primitives: (1*) $\textsc{pos} \subseteq \textsc{rep}$, where $\textsc{pos} = \{\mathbf{t}, \mathbf{r}, \mathbf{b}, \mathbf{l}\}$, and $\textsc{orient} \subseteq \textsc{rep}$, where $\textsc{orient} = \{\mathbf{0}, \mathbf{45}, \mathbf{90}, \mathbf{135}\}$.

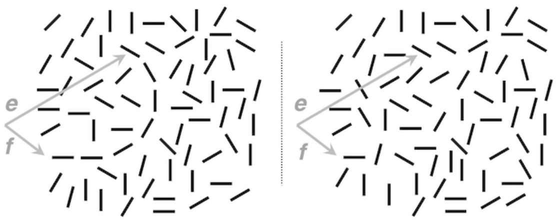
(2*) If $\rho \in \textsc{pos}$ is a representation of x in δ, then it is perfectly accurate iff x is located at position $\mu(\rho)$ in δ.

(3*) If $\theta \in \textsc{orient}$ is a representation of x in δ, then it is perfectly accurate iff x has orientation $\mu(\theta)^\circ$ in δ.

In principle, many such mappings, μ, are possible. For example, the mapping between psychological and objective magnitudes is rarely linear, and may be one of any variety

---

[11] 'Analog representation' is sometimes taken to mean a representation that is continuously rather than discretely variable (see Maley, [2010]). While I assume that there are discretely many types of representational variable or dimension (for representing size, orientation, and so on), it is not essential that there be discretely many representations within a given dimension or variable.



(a) (b)

Figure 3: The two line segments, *e* and *f*, in (a) appear to be part of a common contour, while an identical pair of line segments in t(b) does not. Based on (Geisler and Super, [2000], p. 682).

of logarithmic or exponential functions (Beck, [2019]). It is an empirical matter what mapping yields empirically adequate accuracy conditions for a perceiver at a time.

Many have argued that perceptual representations are analog in something like the sense explained above. The present discussion extends these accounts, showing how analog primitives can compose into structured representations.

4 From Fragments to Contours: Contour Composition

The visual system rarely represents mere arrays of independent contour segments. Contour elements that bound small portions of objects are of limited ecological and psychological significance considered on their own. The value of representing segments of contour lies in the ability to combine such representations into representations of cohesive boundaries of objects, which in turn carry information about the object's shape and kind.

To take a still simplified case: one likely perceives the elements *e* and *f* as part of a common, extended contour in Figure 3a, but not when viewing an identical pair of elements in Figure 3b. Suppose there is no difference in how one represents the orientation, position, and other features of the elements themselves. A semantics of contour representations must explain how contents about coherent, extended contours arise and how they relate to contents concerning the component contour elements. I will argue that neither concatenation nor feature composition are adequate for such an account. Rather, we must posit a distinct mode of combination, 'contour composition'. I begin by arguing that most representations of extended contours are not atomic.

4.1 Coding Contours

One possibility is that extended contours are represented only atomically—they are like representations of smaller contour elements, but on a larger scale. In representing a scene, the visual system merely concatenates atomic representations of contours of different scales. Psychological theories of contour perception reject this possibility and instead hold that the bounding contours of objects normally are 'represented piecemeal' and that 'the visual system assembles these pieces into the coherent percepts of whole objects we experience' (Elder, [2015], p. 207).

Two patterns of evidence suggest that representations of extended contours normally are structured. First, evidence suggests that when representing a coherent contour, the representations of the different segments are not merely concatenated with the rest of the scene. Consider studies of 'grouping advantages'. Sets of elements that are perceived as belonging to the same contour, such as the elements making up the contour in Figure 3a, are more visually salient and more quickly, accurately, and precisely detected than sets of elements that are not perceived as unified (Kapadia et al., [1995]; R. J. Green et al., [2018]). Likewise, one normally is more effective at shifting attention within parts of a common whole (Barenholtz and Feldman, [2003]) and at detecting odd-ones-out among those parts (Kempgens et al., [2013]). These advantages for cohesive contours are not merely effects of recognition, familiarity, or training, since advantages can obtain for entirely unfamiliar stimuli.

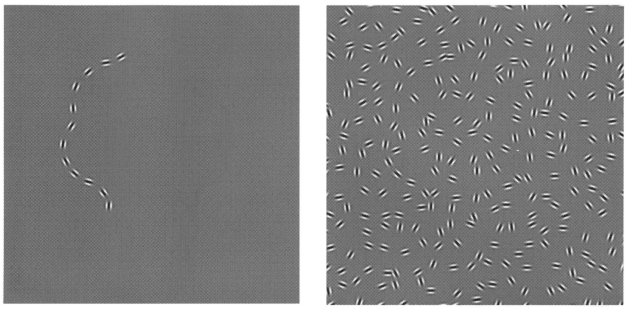
The classic explanation of grouping advantages is that the representations of the grouped elements are constituents of a common representation, which does not include representations of other background elements (Palmer, [1977]). Grouping advantages arise because it is easier and more efficient to encode, store, and manipulate sets of representations that are parts of a common parent representation. It is also easier and more efficient to query relationships between sibling constituents of a common parent representation, since the search space is restricted to the constituents of that representation. To the extent that we represent the individual segments, it appears that these representations are combined together without being concatenated with representations of the rest of the scene. Plausibly, the representation of an extended contour just consists in such selective combination of representations of segments.

An alternative possibility is that the representations of the segments exhibit a grouping advantage because they are associated with a distinct atomic representation of an extended contour that spatially overlaps those segments. Perhaps the processing of the segments is enhanced by this association. However, the representation of a whole contour often requires representing the features of component segments of that contour and encoding their relations. If representations of the segments were merely associated with an atomic representation of the whole contour, it is unclear why the latter would necessarily depend on the former. The standard explanation instead is that the representation of the whole contour requires the representations of the segments because the latter are constituents of the former (Lande, [2021a], pp. 653–56).

For example, Field et al. ([1993]) studied the perception of contours using a 'path detection' paradigm. In a path detection task, subjects must detect the presence of an extended contour or path, such as **??**, in a stimulus such as **??**. These stimuli are constructed by sampling discrete Gabor patches from a smooth contour, which are then embedded in an array of other randomly oriented patches. These Gabor patches enable one to systematically control the spatial information available in the stimulus. In path detection stimuli, there are no coarse spatial properties that could be used to segregate the path from the background. Rather, segregating the path requires representing the positions and orientations of the segments. Such stimuli make it practically impossible to atomically represent an extended contour independently of its component elements.

In such stimuli, grouping advantages for the segments cannot be due to enhancement from some independent atomic representation of the whole path. Rather, the best explanation is that grouping advantages here occur because the representations of the segments combine into a compound representation of the whole path.

Even in more naturalistic stimuli with greater spatial information and connected ele-



(a) (b)

Figure 4: The contour 'path' on the left can be detected in the midst of distractors as in (b), even when there are no coarse spatial properties to differentiate the path from the distractors. Adapted from (Field et al., [1993], p. 177) [reprinted with permission].

ments, the representation of specific kinds of segments along the contour—for example, segments marked by having similar curvature (Baker et al., [2020]) or convexity (Singh, [2015])—seems to be integral to representing the whole contour. These considerations suggest that extended contours normally are not represented atomically and concatenated with representations of other elements of the scene, but are in fact coded in terms of privileged kinds of segments.[12]

While there are various proposed schemes for how contour representations are structured, as a first approximation we can draw on Elder and Goldberg ([2002])'s model, in which contour representations are ascribed the structure of ordered sequences. Cast within our semantic framework, the model says that contour representations have the structure, $\ulcorner(\alpha_1, \ldots, \alpha_n)\urcorner$, where α_i are representations of contour segments (that is, $\alpha_i \in \textsc{contour}$).

This model accommodates the two considerations discussed above. First, a combination of representations of segments, $\ulcorner\{(\alpha_1, \ldots, \alpha_n), \ldots\}\urcorner$, exhibits a grouping advantage that would not arise if the contour elements were merely concatenated together with background elements, $\ulcorner\{\alpha_1, \ldots, \alpha_n, \ldots\}\urcorner$. Second, one cannot represent a contour without representing its segments because a representation of the form, $\ulcorner(\alpha_1, \ldots, \alpha_n)\urcorner$, cannot be tokened without tokening the constituent representations, $\alpha_1, \ldots, \alpha_n$.

4.2 Constraints on Contour Composition

Extended contours are rarely extracted as unitary chunks; they normally are represented by packaging together representations of segments. But the boundaries of objects can be irregular, they can overlap, intersect, and interrupt each other. The visual system faces the challenge of determining which segments go together to bound a common object and which do not. By 'contour composition', I mean the way in which contour representations are related such that they are constituents of a compound contour representation. (I reserve the term 'contour integration' for the process that forms these compound representations.)

---

[12] While attention to component segments of the path can modulate contour integration, it is not thought to be required (Driver et al., [2001]). For the whole path to be a target of attention, the segments must be represented (even if preattentively), and those representations must be combined.

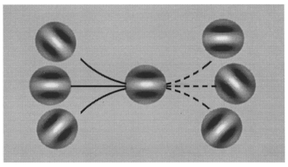


Figure 5: The ‘association field’ is a model that specifies which pairs of contour representations can be combined (solid lines) and which cannot (broken lines). From (Field et al., [1993], p. 190) [reprinted with permission].

Whereas there are few restrictions on how representations can be concatenated, there are more substantive constraints on contour composition. Such constraints function to explain, other things equal, which contour representations are possible as a function of how the constituent representations of segments are related.

Field et al.’s path detection experiments showed that when the represented positions of elements in a ‘path’ are too far apart or at too great an angle from each other, the path is not detectable. They developed the ‘association field’ model, illustrated in Figure 5, to specify which pairs of representations of elements can be combined into a representation of a coherent path and which cannot. This model offers a precise, parameterized definition of the Gestalt rule of ‘good continuation’, which states that for a set of elements to be seen as a cohesive contour, their represented positions and orientations must approximately conform to a smooth curve. The model generalizes beyond artificial path detection stimuli, explaining our abilities to represent contours that are fully connected and naturalistic (Geisler, Perry et al., [2001]; Elder and Goldberg, [2002]), extended in depth (Hess and Field, [1995]; Hess, Hayes et al., [1997]; Khuu et al., [2016]), partially occluded or camouflaged (Kellman, Garrigan and Shipley, [2005]), and which contain corners (Persike and Meinhardt, [2016]).[13]

As stated, the association field constitutes a constraint on the pairwise combination of contour representations. But most contours are coded in terms of more than two segments. On the hypothesis that compound contour representations have the structure of an ordered sequence (Elder and Goldberg, [2002]), so that a pair of representations can be ‘adjacent’ or ‘neighboring’ constituents in the sequence, one can predict contour detection performance with the following rule: an arbitrarily long sequence of contour representations is combinable insofar as each adjacent pair of constituents in the sequence is combinable.[14]

---

[13] Tyler Burge ([2022], p. 82–90), in arguing that contour integration is not a constancy operation, notes that many models of contour integration are concerned with explaining the registration of retinal silhouettes rather than the representation of distal object contours. In fact, many studies of contour integration employ two-dimensional, frontoparallel stimuli and so do not tease these apart. The evidence indicates that similar principles govern the integration of pre-constancy registrations of retinal edges and the combination of post-constancy representations of object contours. As with retinal silhouettes, complex object boundaries normally do not receive atomic representations but are coded in terms of certain types of fragments. And the detection of three-dimensional contours depends on good continuation in the represented positions and orientations of fragments in depth, not just in the image plane. Nothing here hinges on whether contour integration is itself a constancy operation. Perceptual grouping principles are often iterated across pre-constancy and post-constancy processes (Palmer et al., [2003]).

[14] More precisely, Elder and Goldberg ([2002]) model contour representations as Markov chains, in which the combinability of a representation with its successor in the sequence is independent of its combinability with its predecessor. Violations of the Markov assumption are common (Elder, [2015]) and relate

Let '$\circ$' be a contour operator, which takes representations of contours, $\alpha_1, \ldots, \alpha_n \in$ CONTOUR, and returns compound contour representations, $\ulcorner(\alpha_1, \ldots, \alpha_\mathrm{n})\urcorner \in$ CONTOUR:[15]

> Contour Composition (SYN): $\ulcorner\alpha_1 \circ \ldots \circ \alpha_\mathrm{n}\urcorner \in$ CONTOUR, where $\ulcorner\alpha_1 \circ \ldots \circ \alpha_\mathrm{n}\urcorner = \ulcorner(\alpha_1, \ldots, \alpha_\mathrm{n})\urcorner$, iff for every constituent $\alpha_{i:1\leq i\leq(n-1)}$:
>
> (a) $\alpha_i \in$ CONTOUR, and
>
> (b) $\ulcorner\alpha_i \circ \alpha_{i+1}\urcorner \in$ CONTOUR, where
>
> (c) $\ulcorner\alpha_i \circ \alpha_{i+1}\urcorner \in$ CONTOUR iff $align(\alpha_i, \alpha_{i+1})$.

This indicates that contour representations have the structure of an ordered sequence. (a) says that the constituents are themselves contour representations. (b) says that a sequence of representations can be integrated just in case each adjacent pair in the sequence is combinable. And (c) says that each pair is combinable just in case they are well-aligned (that is, they are within each other's association field).[16]

Good continuation is not simply a rule of contour 'inference', so to speak; it is an aspect of the 'grammar' of contours. The principle systematically determines the possibility of a compound contour representation as a function of how its constituents are related. The constraints on contour composition, moreover, differ from those of concatenation and feature composition.

4.3 The Semantics of Contour Composition

Contour composition has distinctive semantic import. It is not enough that the represented contour and its segments all be represented accurately, as with concatenation. It is not necessary that they be identical, as with feature composition. For the compound contour representation to be accurate, the contour segments must be parts of the same contour. This is a natural interpretation of the function of packaging together representations of segments. The structural constraint of good continuation helps to determine which sets of segments bound a common object and closely matches the likelihood that the represented segments are parts of the same object (Sigman et al., [2001]; Geisler, Perry et al., [2001]; Elder and Goldberg, [2002]; Yang and Purves, [2003]).

It does not follow that compound contour representations contain a separate representation, '**partof**()' say, that is dedicated to representing the relevant parthood relation. The simplest hypothesis is that parthood content is built into contour composition:

> Contour Composition (SEM): If $\ulcorner\alpha_1 \circ \ldots \circ \alpha_\mathrm{n}\urcorner \in$ CONTOUR is a representation of x in δ, then it is perfectly accurate iff x at least partially bounds an object in δ, and for each constituent, $\alpha_{i:1\leq i\leq n}$:

---

to the perception of emergent features and configural effects, to be discussed in Section 5.

[15] See also the 'grouping operator' in (Geisler and Super, [2000]) and (Geisler, Perry et al., [2001]).

[16] Some models of contour integration are binary: either a set of contour representations can compose or they cannot (Field et al., [1993]; Kellman and Fuchser, [forthcoming]). By contrast, probabilistic theories of contour integration model the degree of combinability of two contour representations as a function of the likelihood that those two representations would occur as constituents of an integrated contour representation versus the likelihood that they would occur unrelated to each other (Elder and Goldberg, [2002]; Geisler, Perry et al., [2001]; Feldman et al., [2014]). This relative likelihood, or 'binding strength', reflects a generative model of how contour representations decompose. For simplicity's sake, I will not get into the probabilistic versions of these principles.

(a) α_i is perfectly accurate, and

(b) if there exists a y such that α_i is a representation of y in δ, then y is a part of x in δ.

We can now account for the way in which *e* and *f* are perceived differently in the context of Figure 3a than in the context of Figure 3b. Suppose that *e* and *f* receive the same atomic representations when one views either panel. In Fig. 3a, these representations are constituents of a compound contour representation, which represents *e* and *f* as parts of a common contour. Those same atomic representations cannot be combined in one's representation of Fig. 3b, and therefore are not represented as parts of a common contour.

We can moreover justify the assumption that atomic representations of isolated elements, such as those depicted in Figure 2, are representations as of object contours even when the rest of the object has been masked. An account of the content of a representation must accommodate the way that the representation contributes to the contents of compound representations of which it is a part. The representations of isolated elements are of little psychological value in themselves; their basic function is to enter into contour composition. Contour composition functions to represent whole object contours, the parts of which are represented by its constituents. So to fulfill their role in contour composition, the constituents must function to represent parts of object contours.

5 Compositionality and Holism

The Gestalt psychologists famously emphasized that one does not perceive the world as an array of independent elements. They emphasized the 'holistic' character of perceptual organization, by which we perceive elements of a scene as parts of whole objects or configurations. There are two central signatures of holistic or 'Gestalt', phenomena in perceptual organization. First, we often represent 'emergent features': we see wholes as having global features that their parts lack, and we see the parts as having certain features in virtue of our seeing them as parts of a configuration. Second, 'configural effects' obtain in which the way we represent some aspect of a stimulus varies depending on how we represent other aspects of the stimulus.

One might be skeptical that the kind of semantic theory sketched above can go beyond simple cases of contour perception to account for such holistic phenomena. How can the contents of perceptual representations be composed from the contents of their constituents when, as Kurt Koffka ([1936], p. 176) insisted, 'the whole is something else than the sum of the parts'? I will argue that both the representation of emergent features and configural effects are present in contour perception. Yet the semantic theory developed above can in fact help to explain these phenomena. Therefore, there is no incompatibility, in principle, between the holistic character of perception and its semantic compositionality.

5.1 Emergent Features

To a first approximation, emergent features are features that are perceptually attributed to collections of represented items (contour closure, symmetry, and so on), or which are attributed to individual items in virtue of their relationship to other items in a collection (being an odd-one-out) (Pomerantz and Cragin, [2015]). How can one represent these emergent features by simply combining representations of non-emergent features?

While concatenation does not yield content about emergent features, feature and contour composition do. Feature composition has the import not only that the constituent feature representations are accurate, but that they are of the same item in the scene. Contour composition has the import that the constituent representations are of parts of a common contour. The constituents themselves, individually or jointly, do not have such import.

One might well extend the account of the semantics of contour composition to incorporate the composition of content about other emergent features such as symmetry, degree of closure, global size, and so on, as a function of the local features represented by constituents. The general point now is just that content about emergent features can be introduced as a function of how constituent representations are combined.

5.2 Configural Effects

Configural effects, or context effects, are closely related to the representation of emergent features, and occur when the way one represents one aspect of a scene (something's orientation, size, shape, colour, motion, location, and so on) varies in relation to the way one represents other aspects of the scene. Such effects are widespread in perception.

Some have interpreted analogous linguistic phenomena—in which the interpretation of an expression seems to depend on the context in which the expression is embedded—as evidence against strict compositionality in language. Jaako Hintikka ([1983], p. 265), for example, wrote that 'Whenever the meaning (interpretation) of an expression depends on the wider context in which it is embedded, a violation of compositionality is in the offing'. In the same vein, some philosophers have worried that configural effects undermine the prospects for a systematic semantics of perception (Cummins, [1996]; Camp, [2018]).

However, we have already encountered and explained one configural effect in contour perception: one sees e and f as parts of the same contour in the context of Figure 3a but not in the context of Figure 3b. The perceived unity of these elements seems to depend on the presence and arrangement of other elements that one sees—what the Gestalt theorist Max Wertheimer ([1938]) called 'pro-structural' elements.

Contour composition accounts for this configural effect. The representations of e and f in Figure 3 are not in themselves pairwise combinable through contour composition—they do not fall within each other's association fields. Two representations that are not pairwise combinable can only be combined as parts of a longer sequence of representations in which adjacent pairs can combine. The elements in Fig. 3a, but not Fig. 3b, are arranged so that one can form a sequence of representations that link the representation of e to the representation of f. Insofar as the representations of e and f compose together in Fig. 3a, one represents e and f as part of the same contour. Insofar as e and f can only be concatenated in Fig. 3b, one does not represent them as parts of the same contour.

Whereas e and f require 'pro-structural' elements in order to be represented as parts of the same whole in Figure 3, in other cases the introduction of 'contra-structural' elements can destroy the represented unity of a set of elements. For example, the three line segments that make up the 'U'-shaped part in Figure 6 are represented as parts of a common, complete part of a figure in panels H, MH, and M, but not in panels ML and L (Palmer, [1977]). In the latter cases, even if the representations of the three line segments

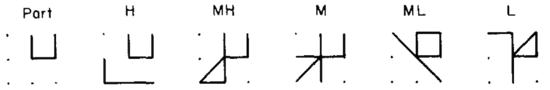


Figure 6: The set of contours that define the 'U'-shaped part are represented as a unified, complete part in panels H, MH, and M, but not in ML and L. Adapted from (Palmer, [1977]) [reprinted with permission].

could in principle combine together, the preferred structures are ones in which the representations of those segments either do not all combine together (as in L), or in which they combine with further representations (as in ML). These sorts of contra-structural effects can be explained by positing global constraints (Kellman and Fuchser, [forthcoming]) or priors favoring representations that are structurally simpler (Feldman et al., [2014]) or that represent certain emergent features such as closure (Elder, [2015]).

As Schwartz et al. ([2007], p. 530) note, in some cases the explanation of configural effects may call for reconsidering the semantics of certain representations. For example, the way the orientation of a contour is represented depends on the perceived orientation of its surround. It is easier to detect a tilted line segment among vertical distractors (Figure 7a) than it is to detect a vertical line segment among tilted distractors—a 'visual search asymmetry' (Treisman and Gormican, [1988]). All things equal, tilted elements are more visually salient than vertical elements. However, if the same array of elements are surrounded by a frame that is tilted in the same direction as the tilted line segment (Figure 7b), the search asymmetry reverses: now it is the tilted element that is less salient and harder to detect among distractors (May and Zhaoping, [2009]; Marendaz, [1998]).

It is implausible that the surrounding frame simply induces a substantial illusion as to the orientations of the surrounded elements. Instead, a standard explanation for the reversal is that the line segment is not simply represented as 'tilted' or 'vertical' full-stop, where the accuracy of such orientation representations is evaluated with respect to some fixed reference such as the direction of gravity. Rather, the visual system represents elements as having relative orientations—as 'tilted relative to x' or 'vertical relative to x', where the orientation of x is supplied by the representation of the surrounding frame's orientation. The objective orientation of the element relative to the scene is represented compositionally in terms of the orientation of the element relative to the surrounding frame together with the orientation of the frame in the scene (Rock, [1990]). To accurately represent an objectively tilted element would require assigning different relative orientations depending on the represented orientation of the element's surroundings, while accuracy would demand representing both a tilted element in a tilted frame and a vertical element in an upright frame as having the same relative orientations.

Configural effects arise in simple cases of contour and form perception. These cases illustrate how compositional analyses of configural effects may proceed either by identifying differences in how representations are structured across configurations, or by refining our understanding of the semantics of the representations involved. There is therefore no *a priori* tension between configural effects and compositional analysis.[17]

---

[17] Analogous points have been made in the context of explaining sentential context effects within a compositional theory of language (Janssen, [2011]).

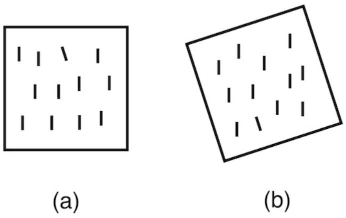


Figure 7: (a) When surrounded by an upright frame, it is easier to find a tilted bar among vertical distractors than a vertical bar among tilted distractors. (b) When elements are surrounded by a tilted frame, the search asymmetry reverses.

6 The Format of Contour Representations

I have argued that representations of contours are composed from discrete representations of contour segments, which in turn are composed from distinct representations of features of those segments. However, this does not mean that perceptual representations are 'discursive' or 'language-like'. Many types of structured icons, such as maps (Camp, [2007]; Rescorla, [2009]) and diagrams (Shin, [1994]), are semantically compositional without being substantively language-like. Contour representations exhibit a hallmark feature of such structured icons, which I will call 'domain-specific form': as a first approximation, structural relations within and between these representations mirror substantive relations in the represented domain (Barwise and Etchemendy, [1996]; Burge, [2022]).[18]

Consider spatial 'heat maps', which use gradations of color to visualize magnitudes along one dimension (population density, say) in relation to spatial layout (electoral districts, say). First, whereas relations between primitive symbols of predicate calculus do not carry any significance about relations among their semantic values, the primitives of heat maps are analog: darker gradations of patches in a heat map typically signify greater magnitudes on the represented dimension. Second, structural relations between different marks on the map function to correspond to spatial relations among the represented regions. Whereas the syntactic operation of applying a predicate to a term may correspond to the metaphysical relation of a property being instantiated in an individual, this correspondence is highly abstract and domain-general in comparison to the spatial and topological commitments embedded in the structure of a map (Camp, [2007], p. 157). Third, the constraints on combining elements within a map guarantee some level of spatial and topological consistency in the map's content, whereas a formula's well-formedness in predicate calculus does not even guarantee its logical consistency (Shimojima, [2001]).

Contour representations have domain-specific form of the sort that paradigmatically characterizes structured icons such as maps and diagrams. In the first place, the primitive representations of features such as orientation, position, size, and so on are analog. Psychological similarity relations among such representations function to correspond to certain objective similarity relations among the represented features. The accuracy conditions of a primitive representation are partly a function of how that representation falls

---

[18] I am not defending a constitutive account of what makes something a map or a diagram, nor am I arguing that contour representations just are a type of map or diagram. My aim is only to highlight a significant feature of the format of contour representations.

under the relevant psychological relations. Moreover, whether two contour representations can combine is a function of how psychologically similar their respective positions and orientations are. So both the content and the combinatorial potential of a contour representation are a function of its psychological similarity to other contour representations.

In the second place, the structure of a given contour representations (at least down to the atoms) mirrors, and carries semantic import about, the part-whole structure and geometry of the represented contour. For $\mathbf{e_1} \circ \ldots \circ \mathbf{e_n}$ to be an accurate representation of a contour x, its constituents, $\mathbf{e_1}, \ldots, \mathbf{e_n}$, must be accurate representations of parts of x. Contour representations conform to something like the 'Parts Principle', which holds that the parts of a picture represent parts of what the picture represents (Sober, [1976]; Fodor, [2007]). Moreover, the linear ordering of the contour representation corresponds to the linear ordering of represented contour segments. Structurally adjacent constituents function to represent topologically neighboring parts of the represented contour.

Finally, the very combinability of contour representations entails the good continuation of their represented positions and orientations. This principle of combination is not topic-neutral; it carries a semantic commitment about the shapes of represented contours. This commitment reflects the statistics of contours in natural scenes: smoothly aligned contours are much more likely to be part of the same object than unaligned contours.

These features—analog primitives, representational structure that mirrors represented structure, and combinatorial constraints that reflect constraints in the represented domain—often are taken to be hallmarks of iconic representations. At the same time, contour representations do not exhibit other features that have been taken to be characteristic of icons as opposed to formulas. For example, it is commonly held that pictures lack canonical constituent structure (Fodor, [2007]) and that icons represent multiple features in a 'holistic' way such that there is no distinction between the part of an icon that represents an item's shape and the part of the icon that represents its colour (E. Green and Quilty-Dunn, [2017]). By contrast, the theory outlined here gives canonical analyses of contour representations, on which an item's position and orientation receive separate representations (though neither can occur in isolation).

The present account entails that contour representations have determinate compositional structure. But like maps and diagrams, and unlike the semantically bare and topic-neutral principles of composition that characterize logic and arguably much of language, the structure and combinatorial constraints of contour representations are semantically rich and embed certain commitments about the represented domain of object contours.

7 Conclusion

Semantic theories of mental representations can illuminate the ways that mental states fit together to represent complex situations. They can in turn inform our understanding of the computational, epistemic, and constitutive relations between mental states. I have tried to contribute to this end by giving a systematic, empirically grounded semantics of contour representations in mid-level vision. I have argued that visual representations of object contours in a scene are constructed according to several different modes of composition—including feature composition, contour composition, and concatenation—and I have given an account of their structural constraints and semantic import.

This provisional account has several principal payoffs. First, it models a more general working framework for characterizing the structure and content of perceptual representations. I have highlighted a core explanatory role of representational accuracy conditions as well as a distinct explanatory role of representational structure. Other things equal, a representation's accuracy conditions determine how well a subject's task responses will track a specific aspect of the distal stimulus, as a function of how accurate the representation is. Other things equal, structural constraints explain which representations of a given kind are possible in a system and how they can co-vary. Semantic composition principles, which determine the accuracy conditions of a complex representation as a function of the accuracy conditions of its constituents and their mode of combination, are constrained by both of these explanatory roles.

Second, the account demonstrates how a compositional semantics of perception can be compatible with, and even help to explain, the holistic or Gestalt character of perception. Emergent perceptual content and configural effects have often been taken to show that 'the whole is other than the sum of its parts'. Nevertheless, such phenomena are consistent with the idea that much of what we perceive is coded in terms of more elementary parts and features under specific modes of combination.

Finally, the account illuminates the format of an important type of visual representation. Visual contour representations exhibit a hallmark feature of structured icons such as diagrams and maps, while differing in that respect from typical discursive representations of logic and language. Namely, structured representations of contours have domain-specific form, which reflects substantive commitments about the nature of object contours.

The present study contributes to the the development of a semantics of perception. But contour representations make up only a fragment of our psychological system. It remains to be seen how far this semantic program can be pushed and how far it will need to be extended and revised in order to cover a broader range of perceptual capacities and phenomena, both within vision and across other modalities.

Acknowledgements

I have received invaluable feedback from Jacob Beck, Denis Buehler, Tyler Burge, Rosa Cao, Sam Clarke, Sam Cumming, Gabe Dupre, James Elder, Gabe Greenberg, Gabbrielle Johnson, Phil Kellman, William Kowalsky, Jessie Munton, Bence Nanay, and my anonymous reviewers. Thanks also to audiences at the MindWork workshop at the University of Texas-Austin and the Formats of Vision & Thought workshop in Toronto. This work was supported by funding from an Insight Development Grant from the Social Sciences and Humanities Research Council and by the Canada First Research Excellence Fund through York University's 'Vision: Science to Applications' program.

Department of Philosophy,
and Centre for Vision Research
York University
Toronto, Canada
lande@yorku.ca

References

Algom, D. and D. Fitousi [2016]: 'Half a Century of Research on Garner Interference and the Separability-Integrality Distinction', *Psychological Bulletin*, **142**, pp. 1352–83.

Baker, N., P. Garrigan and P. J. Kellman [2020]: 'Constant Curvature Segments As Building Blocks of 2D Shape Representation', *Journal of Experimental Psychology: General*.

Barenholtz, E. and J. Feldman [2003]: 'Visual Comparisons Within and Between Object Parts: Evidence for a Single-Part Superiority Effect', *Vision Research*, **43**, pp. 1655–66.

Barwise, J. and J. Etchemendy [1996]: 'Heterogeneous Logic', in G. Allwein and J. Barwise (*eds.*), *Logical Reasoning with Diagrams*, New York: Oxford University Press, pp. 179–200.

Beck, J. [2019]: 'Perception is Analog: The Argument From Weber's Law', *The Journal of Philosophy*, **116**, pp. 319–249.

Biederman, I. and G. Ju [1988]: 'Surface Versus Edge-Based Determinants of Visual Recognition', *Cognitive Psychology*, **20**, pp. 38–64.

Block, N. [2023]: *The Border between Seeing and Thinking*, New York: Oxford University Press.

Burge, T. [2010]: *Origins of Objectivity*, Oxford: Oxford University Press.

Burge, T. [2022]: *Perception: First Form of Mind*, Oxford: Oxford University Press.

Camp, E. [2007]: 'Thinking with Maps', *Philosophical Perspectives*, **21**, pp. 145–82.

Camp, E. [2018]: 'Why Maps Are Not Propositional', in A. Grzankowski and M. Montague (*eds.*), *Non-Propositional Intentionality*, Oxford: Oxford University Press, pp. 19–45.

Casati, R. and A. C. Varzi [1999]: *Parts and Places: The Structures of Spatial Representation*, The MIT Press.

Christensen, J. H., P. J. Bex and J. Fiser [2019]: 'Coding of Low-Level Position and Orientation Information in Human Naturalistic Vision', *PLOS ONE*, **14**, e0212141.

Clark, A. [2004]: 'Feature-Placing and Proto-Objects', *Philosophical Psychology*, **17**, pp. 443–69.

Clarke, S. [2021]: 'Mapping the Visual Icon', *The Philosophical Quarterly*.

Cummins, R. [1996]: 'Systematicity', *The Journal of Philosophy*, **93**, pp. 591–614.

Davies, W. [2020]: 'Colour Relations in Form', *Philosophy and Phenomenological Research*.

Driver, J., G. Davis, C. Russell, M. Turatto and E. Freeman [2001]: 'Segmentation, Attention and Phenomenal Visual Objects', *Cognition*, **80**, pp. 61–95.

Elder, J. H. [2015]: 'Bridging the Dimensional Gap: Perceptual Organization of Contour Into Two-Dimensional Shape', in J. Wagemans (*ed.*), *The Oxford Handbook of Perceptual Organization*, Oxford University Press, pp. 207–35.

Elder, J. H. and R. M. Goldberg [2002]: 'Ecological Statistics of Gestalt Laws for the Perceptual Organization of Contours', *Journal of Vision*, **2**, pp. 324–53.

Feldman, J. [1999]: 'Does Vision Work? Towards a Semantics of Perception', in E. Lepore and Z. Pylyshyn (*eds.*), *What is Cognitive Science?*, Basil Blackwell, pp. 208–29.

Feldman, J., M. Singh and V. Froyen [2014]: 'Perceptual Grouping as Bayesian Mixture Estimation', in S. Gepshtein, L. Maloney and M. Singh (*eds.*), *The Oxford Handbook of Computational Perceptual Organization*, Oxford University Press.

Field, D. J., A. Hayes and R. F. Hess [1993]: ‘Contour Integration by the Human Visual System: Evidence for a Local “Association Field”’, *Vision Research*, **33**, pp. 173–93.

Fodor, J. [1975]: *The Language of Thought*, Cambridge, MA: Harvard University Press.

Fodor, J. [2007]: ‘The Revenge of the Given’, in B. P. McLaughlin and J. D. Cohen (*eds.*), *Contemporary Debates in Philosophy of Mind*, Blackwell, pp. 105–16.

Fodor, J. and Z. Pylyshyn [1988]: ‘Connectionism and Cognitive Architecture: A Critical Analysis’, *Cognition*, **28**, pp. 3–71.

Garner, W. R. [1974]: *The Processing of Information and Structure*, New York: Lawrence Elbaum Associates.

Gauker, C. [2012]: ‘Perception Without Propositions’, *Philosophical Perspectives*, **26**, pp. 19–50.

Geisler, W. S., J. Perry, B. Super and D. Gallogly [2001]: ‘Edge Co-Occurrence in Natural Images Predicts Contour Grouping Performance’, *Vision Research*, **41**, pp. 711–24.

Geisler, W. S. and B. J. Super [2000]: ‘Perceptual Organization of Two-Dimensional Patterns’, *Psychological Review*, **107**, pp. 677–708.

Green, E. and J. Quilty-Dunn [2017]: ‘What is an Object File?’, *British Journal for the Philosophy of Science*, **72**, pp. 665–99.

Green, R. J., J. E. Dickinson and D. R. Badcock [2018]: ‘Integration of Shape Information Occurs Around Closed Contours But Not Across Them’, *Journal of Vision*, **18**, pp. 1–13.

Heim, I. and A. Kratzer [Jan. 1998]: *Semantics in Generative Grammar*, Wiley.

Hess, R. F. and D. J. Field [1995]: ‘Contour Integration Across Depth’, *Vision Research*, **35**, pp. 1699–711.

Hess, R. F., A. Hayes and F. A. Kingdom [1997]: ‘Integrating Contours Within and Through Depth’, *Vision Research*, **37**, pp. 691–6.

Hintikka, J. [1983]: ‘Theories of Truth and Learnable Languages’, in *The Game of Language: Studies in Game-Theoretical Semantics and Its Applications*, Boston: D. Reidel Publishing Company, pp. 259–92.

Hubel, D. H. and T. N. Wiesel [1968]: ‘Receptive Fields and Functional Architecture of Monkey Striate Cortex’, *The Journal of Physiology*, **195**, pp. 215–43.

Janssen, T. M. [2011]: ‘Compositionality’, in J. van Benthem and A. ter Meulen (*eds.*), *Handbook of Logic and Language*, 2nd, Elsevier, pp. 495–553.

Kapadia, M. K., M. Ito, C. D. Gilbert and G. Westheimer [1995]: ‘Improvement in Visual Sensitivity by Changes in Local Context: Parallel Studies in Human Observers and in V1 of Alert Monkeys’, *Neuron*, **15**, pp. 843–56.

Keane, B. P. [2018]: ‘Contour Interpolation: A Case Study in Modularity of Mind’, *Cognition*, **174**, pp. 1–18.

Kellman, P. J. and V. Fuchser [forthcoming]: ‘Visual Completion and Intermediate Representations in Object Formation’, in A. Mroczko-Wąsowicz and R. Grush (*eds.*), *Sensory Individuals: Unimodal and Multimodal Perspectives*, Oxford University Press.

Kellman, P. J., P. Garrigan and G. Erlikhman [2013]: ‘Challenges in Understanding Visual Shape Perception and Representation: Bridging Subsymbolic and Symbolic Coding’, in S. Dickinson and Z. Pizlo (*eds.*), *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective*, London: Springer-Verlag, pp. 249–74.

Kellman, P. J., P. Garrigan and T. F. Shipley [2005]: ‘Object Interpolation in Three Dimensions’, *Psychological Review*, **112**, pp. 586–609.

Kempgens, C., G. Loffler and H. S. Orbach [2013]: ‘Set-Size Effects for Sampled Shapes: Experiments and Model’, *Frontiers in Computational Neuroscience*, **7**, pp. 1–18.

Khuu, S. K., V. Honson and J. Kim [2016]: ‘The Perception of Three-Dimensional Contours and the Effect of Luminance Polarity and Color Change on Their Detection’, *Journal of Vision*, **16**, pp. 1–15.

Kimchi, R. [1992]: ‘Primacy of Wholistic Processing and Global/Local Paradigm: A Critical Review’, *Psychological Bulletin*, **112**, pp. 24–38.

Koenderink, J. J. [1984]: ‘What Does the Occluding Contour Tell Us About Solid Shape?’, *Perception*, **13**, pp. 321–30.

Koffka, K. [1936]: *Principles of Gestalt Psychology*, London: Kegan Paul, Trench, Trubner & Co., Ltd.

Lande, K. J. [2021a]: ‘Mental Structures’, *Noûs*, **55**, pp. 649–77.

Lande, K. J. [2021b]: ‘Seeing and Visual Reference’, *Philosophy and Phenomenological Research*.

Larson, R. and G. Segal [1995]: *Knowledge of Meaning: An Introduction to Semantic Theory*, Cambridge, MA: MIT Press.

Maley, C. J. [2010]: ‘Analog and Digital, Continuous and Discrete’, *Philosophical Studies*, **155**, pp. 117–31.

Marendaz, C. [1998]: ‘Nature and Dynamics of Reference Frames in Visual Search for Orientation: Implications for Early Visual Processing’, *Psychological Science*, **9**, pp. 27–32.

Marr, D. [1982]: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Cambridge, MA: The MIT Press.

Matthen, M. [2005]: *Seeing, Doing, & Knowing: A Philosophical Theory of Sense Perception*, Oxford: Oxford University Press.

May, K. A. and L. Zhaoping [2009]: ‘Effects of Surrounding Frame on Visual Search for Vertical or Tilted Bars’, *Journal of Vision*, **9**, pp. 1–19.

Palmer, S. E. [1977]: ‘Hierarchical Structure in Perceptual Representation’, *Cognitive Psychology*, **9**, pp. 441–74.

Palmer, S. E., J. L. Brooks and R. Nelson [Nov. 2003]: ‘When Does Grouping Happen?’, *Acta Psychologica*, **114**, pp. 311–30.

Pasupathy, A., Y. El-Shamayleh and D. V. Popovkina [2018]: ‘Visual Shape and Object Perception’, in *Oxford Research Encyclopedia of Neuroscience*, http://neuroscience.oxfordre.com/view/ 9780190264086-e-75.

Persike, M. and G. Meinhardt [2016]: ‘Contour Integration with Corners’, *Vision Research*, **127**, pp. 132–40.

Pomerantz, J. R. and A. I. Cragin [2015]: ‘Emergent Features and Feature Combination’, in J. Wagemans (*ed.*), *The Oxford Handbook of Perceptual Organization*, Oxford: Oxford University Press, pp. 88–107.

Portilla, J. and E. P. Simoncelli [2000]: ‘A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients’, *International Journal of Computer Vision*, **40**, pp. 49–70.

Quinlan, P. T. [2003]: ‘Visual Feature Integration Theory: Past, Present, and Future’, *Psychological Bulletin*, **129**, pp. 643–73.

Rescorla, M. [2009]: ‘Cognitive Maps and the Language of Thought’, *British Journal for the Philosophy of Science*, **60**, pp. 377–407.

Rock, I. [1990]: 'The Frame of Reference', in *The Legacy of Solomon Asch: Essays in Cognition and Social Psychology*, Lawrence Elbaum, pp. 243–68.

Sauvan, X. M. [1999]: 'Orientation Constancy in Neurons of Monkey Visual Cortex', *Visual Cognition*, **6**, pp. 43–54.

Schwartz, O., A. Hsu and P. Dayan [2007]: 'Space and Time in Visual Context', *Nature Reviews Neuroscience*, **8**, pp. 522–35.

Shimojima, A. [2001]: 'The Graphic-Linguistic Distinction', *Artificial Intelligence Review*, **15**, pp. 5–27.

Shin, S.-J. [1994]: *The Logical Status of Diagrams*, New York: Cambridge University Press.

Sigman, M., G. A. Cecchi, C. D. Gilbert and M. O. Magnasco [2001]: 'On a Common Circle: Natural Scenes and Gestalt Rules', *Proceedings of the National Academy of Sciences*, **98**, pp. 1935–40.

Singh, M. [2015]: 'Visual Representation of Contour and Shape', in J. Wagemans (*ed.*), *The Oxford Handbook of Perceptual Organization*, Academic Press, pp. 236–56.

Sober, E. [1976]: 'Mental Representations', *Synthese*, **33**, pp. 101–48.

Taylor, J. M. and Y. Xu [2022]: 'Representation of Color, Form, and Their Conjunction Across the Human Ventral Visual Pathway', *NeuroImage*, **251**, p. 118941.

Treisman, A. and G. Gelade [1980]: 'A Feature-Integration Theory of Attention', *Cognitive Psychology*, **12**, pp. 97–136.

Treisman, A. and S. Gormican [1988]: 'Feature Analysis in Early Vision: Evidence From Search Asymmetries', *Psychological Review*, **95**, pp. 15–48.

Treisman, A. and H. Schmidt [1982]: 'Illusory Conjunctions in the Perception of Objects', *Cognitive Psychology*, **14**, pp. 107–41.

Von Eckardt, B. [2012]: 'The Representational Theory of Mind', in *The Cambridge Handbook of Cognitive Science*, Cambridge University Press, pp. 29–49.

Wade, N. J. and M. T. Swanston [2013]: *Visual Perception: An Introduction*, Third, New York: Taylor & Francis Group.

Wertheimer, M. [1938]: 'Laws of Organization in Perceptual Forms', in W. Ellis (*ed.*), *A source book of Gestalt psychology*, London: Routledge & Kegan Paul, pp. 71–88.

Yang, Z. and D. Purves [2003]: 'Image/Source Statistics of Surfaces in Natural Scenes', *Network: Computation in Neural Systems*, **14**, pp. 371–90.