Compositionality and Context in Perception

Kevin J. Lande
Department of Philosophy
Centre for Vision Research
York University

Abstract

A compositional theory of perceptual representations would explain how the accuracy conditions of a given type of perceptual state depend on the contents of constituent perceptual representations and the way those constituents are structurally related. Such a theory would offer a basic framework for understanding the nature, grounds, and epistemic significance of perception. But an adequate semantics of perceptual representations must accommodate the holistic nature of perception. In particular, perception is replete with context effects, in which the way one perceptually represents one aspect of a scene (including the position, size, orientation, shape, color, motion, or even unity of an object) normally depends on how one represents many other aspects of the scene. The ability of existing accounts of the semantics of perception to analyze context effects is at best unclear. Context effects have even been thought to call into question the very feasibility of a systematic semantics of perception. After outlining a compositional semantics for a rudimentary set of percepts, I draw on empirical models from perceptual psychology to show how such a theory must be modified to analyze context effects. Context effects arise from substantive constraints on how perceptual representations can combine and from the different semantic roles that perceptual representations can have. I suggest that context effects are closely tied to the objectivity of perception. They arise from a perceptual grammar that functions to facilitate the composition of reliably accurate representations in an uncertain but structured world.

Word count: Paper $\approx$ 17,160 / Abstract = 240

1 Introduction

Mental states such as beliefs, desires, memories, and perceptions are representational. The visual state I am in when I see the tabletop in front of me represents that tabletop as a yellow rectangular surface and is accurate insofar as that surface *is* yellow and rectangular. Mental representations can be structured too; they can have constituent parts. The visual state that I am in has as constituents my representation of the yellowness of the tabletop and my representation of its rectangularity, which in turn has as constituents my representation of the sides and corners of the tabletop (Lande, 2021). A systematic semantic theory of mental representations would articulate the principles by which a representation's content relates to its structure. In particular, a *compositional* semantics of mental representations would explain the way a mental state represents the world as a function of the way its constituent parts represent the world and how those parts are structurally related. Rather than simply asking about what a given type of mental state represents—for example, whether some perceptual states represent certain kinds of shape properties, whether some perceptual states represent *cause*, and so on—such a theory asks about how the contents of mental states fit together within a broader representational system.

Many of the attempts to outline a systematic semantics of mental representations have focused on aspects of cognition and thought (Fodor, 1975, Devitt, 2006, Camp, 2007, Rescorla, 2009). Yet a growing consensus holds that accounts of the nature of mental representation and of its place in nature must be grounded in an adequate account of perceptual representation (Burge, 2010, Neander, 2017). Perception is arguably the most basic species of mental representation, or intentionality. A semantics of perceptual representations would offer a basic framework for understanding the nature, grounds, and epistemic significance of perception. The aim of this paper is to make progress toward an adequate compositional semantics of perception.

Discussions of the semantics of perception typically focus on simple core cases. For example, as I look at the tabletop, plausibly I see its surface as rectangular, yellow, and on the right, rather than round, blue, and on the left, solely in virtue of the facts that my visual representation of the surface is made up of a representation of its rectangular shape, a representation of its yellow color, and a representation of its location, and that these representations are "integrated" together. When I see a round blue tabletop next to the rectangular yellow tabletop, my representations of these items are in some sense concatenated together as constituents of my total percept. Some characterize these "integration" and "concatenation" operations on the model of conjunction, whether of predicates or of formulas (Sober, 1976, Clark, 2004, Matthen, 2005, Tacca, 2011), while others analogize them to the juxtaposition of marks in an image or map (Fodor, 2008, Kulvicki, 2015, Quilty-Dunn, 2019). On either conception the standard attitude seems to be that the "grammar" of vision—its combinatorial principles and their semantic import—is fairly rudimentary and, as Fodor (2008, p. 175) put it, "unarcane."

We can distill from these views two common assumptions about the compositional structure of perception. First, *semantic uniformity*: a constituent representation of the feature of one part of the scene (for example, the yellowness of the tabletop) has the same semantic role as a constituent representation of any other feature of any other part of the scene (the rectangularity of the tabletop; the blueness of the nearby tabletop). Any such representation can be evaluated on its own as veridical or not with respect to the object being represented. Each determines a "complete" veridicality condition (for example, that the represented item be yellow). Second, *unrestricted combination*: these constituents can in principle be combined without substantive constraints. In principle, the representation of the tabletop's yellowness could just as easily have combined with a representation roundness as with a representation of rectangularity, and a representation of a rectangular yellow tabletop could just as easily combine with a representation of an adjacent round blue tabletop as with a representation of an adjacent black stool. These two assumptions leave open questions about what primitive parts populate one's "perceptual dictionary" and what principles of computation or "inference" govern the formation and use of these parts and their combinations. But even setting these questions to the side, the compositional grammar of perception is unlikely to be as simple as the simple cases suggest.

The Gestalt psychologists of the early 20^{th} century—Kurt Koffka, Max Wertheimer, and Wolfgang Köhler among them—insisted that perception is profoundly holistic. The central mark of the holistic character of perception is the ubiquity of *perceptual context effects*, in which the way one perceives one aspect of a scene varies in relation to the way one perceives other aspects of the scene. For example, the headlights of two motorcycles traveling side by side appear to belong to separate objects during the daytime when the outlines of the motorcycles and the pavement between them are visible. The same pair of lights appear to belong to a single object in the nighttime when these other parts of the scene are invisible. Likewise, the perceived tilt of the floor influences how one perceives the orientation of a picture frame hanging on the wall above. In fact, almost no perceptual phenomena are in principle immune to context effects. How one perceives orientation, size, shape, color, location, motion, unity, and just about any other feature of an object normally depends on how one perceives many other aspects of the scene. Context effects are so pervasive in perception that one can reasonably infer that they are central to its successful functioning.

At best, the ubiquity of context effects calls into question the explanatory power and generality of accounts that focus on cases in which these effects are elided. The Gestalt school targeted the empiricist theories of psychological "structuralists," like Edward Titchener, who held that percepts consist of independent elements that can be associated and combined without restriction. Wertheimer (1938) wrote that the perceived features in a scene are not "pieces to be combined in and-summations." Some today take the ubiquity of perceptual context effects to call into question the very feasibility of a compositional semantics of perception. For example, Camp writes that "it is not obvious that a [compositional] semantics can be offered for pictures, *let alone perception*. In particular, the representational significance of images often appears to be highly local and context-dependent: changes to the marks which make little semantic difference at one location produce significant semantic difference, or destroy the representation altogether, at another" (Camp, 2018, p. 42, my emphasis; see also Cummins, 1996, Balog, 2009). It is natural to wonder whether the project of compositional analysis is fundamentally in tension with the holistic nature of perception.

Focusing on vision, I argue that sophisticated compositional theories can handle characteristically holistic phenomena in perception and illuminate the underlying principles that drive them. I begin, in Section 2, by discussing the explanatory value of a semantics of perception. I sketch a basic semantics for a rudimentary set of percepts, exemplifying the assumptions that the constituents of perceptual representations are semantically uniform and freely combinable. In Section 3, I introduce several classic context effects, which the basic theory as it stands cannot explain. In Section 4, I draw on models in perceptual psychology to develop compositional analyses of these phenomena. These analyses abandon the assumptions of semantic uniformity and unrestricted combination. Context effects arise from substantive constraints on the combination of perceptual representations and from the integration of representations with different semantic roles. In Section 5, I suggest that context effects are closely tied to the objectivity of perception. Context effects function to facilitate the composition of accurate perceptual representations in an uncertain but structured environment. I conclude in Section 6 by discussing the prospects of a semantics of perception.

2 Semantics for Perception

I assume a realist attitude toward semantic theories of mental representations. The value of having a systematic semantics of perception is not just in having a compact way of describing a range of different perceptual states. Mental representations really have content and they really have constituent structure. An adequate semantics of perception must explain how semantic relationships between perceptual states correspond to structural relationships between those states. In this section, I discuss how such a theory would provide a basic framework for addressing questions about the nature, grounds, and epistemic potential of perceptual content. I then sketch a basic semantics for a simplified set of percepts. In the next sections, I will examine how this basic semantics needs to be revised in order to accommodate characteristic perceptual phenomena.

2.1 Compositionality as a Theoretical Framework

I will proceed on the working hypothesis that a systematic semantics of perceptual representations will be compositional. A representational system is *semantically compositional* if and only if the content of every representation in the system is functionally determined solely by the way that representation is structured from its constituents and by the contents of those constituents.[1] A compositional semantic theory identifies the structural principles by which representations can combine and the semantic principles by which the representations with a given structure derive their contents from the contents of their constituents. Truth-functional logic, in which the truth value of every well-formed compound formula or sub-formula is wholly a function of the truth values of its constituents and their mode of combination, is paradigmatically compositional.

Consider the visual state one is in when viewing these three line segments:

— | /

One's visual state represents the stimulus in a certain way. A compositional analysis of this perceptual representation must specify the structure of one's representation of the stimulus. For example: the representation of this scene may have as constituents a representation as of the horizontal line on the left, a representation as of the vertical line in the center, and a representation as of the tilted line on the right. The theory must also specify the semantic import of having a set of constituents under the relevant mode of combination. For example: as a rule this sort of combination is veridical just in case all its constituents are veridical. The analysis entails, plausibly enough, that one's perceptual representation (or "percept") of the stimulus is veridical if and only if the line on the left is horizontal, the one in the middle is vertical, and the one on the right is tilted right.

The central explanatory value of compositional semantic theories lies in their potential to identify and explain systematic relationships between the contents of different representations (Fodor and Pylyshyn, 1988). Consider one's perceptual representation of the following display:

— \ |

How one perceives this arrangement is related to how one perceives the previous one. In both cases, one's percept is veridical only if the line on the left is horizontal. An explanation of this commonality is that both representations have a common constituent that represents a horizontal line on the left. When the constituent figures into a more complex representation, that latter

[1] Though there are various ways to define compositionality more precisely, this formulation will do for our purposes.

representation will be veridical only if there is a horizontal line on the left—assuming that as a rule complex perceptual representations of this form are veridical just in case their constituents are. There is therefore an underlying semantic explanation of why the percepts of the two stimuli above have related veridicality conditions.

Compositional theories can help to identify in the first place which relationships between contents are systematic and which are merely incidental. For example, the percepts of the two stimuli above are veridical only if there is a vertical line segment in the scene, though they locate their respective line segments in different positions. Is it a psychological law that if one of these percepts represents a vertical line segment, then so must the other? One theory might analyze the representations of the two stimuli as both having a constituent representation of a line segment's vertical orientation, separate from a representation of its position. This account entails that the two percepts have related veridicality conditions as a necessary consequence of their having a common type of constituent. An alternative theory might provide independent analyses of the percepts of the two stimuli, taking one to consist of a primitive representation of a vertical line in the middle and the other to consist of a primitive representation of a vertical line on the right. In that case, the content of the one representation does not necessarily imply anything about the content of the other. Evidence for one or the other of these theories constitutes evidence about whether percepts of vertical lines in varying positions make up a common explanatory semantic kind or not.

Compositional theories are apt to identify and explain systematic relationships between the contents of representations because one of the main explanatory roles of representational content in these theories is to encode the unique common contribution that a representation makes in determining the contents of more complex representations of which it is a part. I may find the horizontal line visually pleasing in the second display but ugly in the first. Suppose my representations of both displays contain a common constituent representation of a horizontal line at position 1, which is combined from a representation of the line's orientation and position. It cannot be in virtue of this constituent alone that I represent the line as pleasing in one surround but ugly in another. So this constituent cannot represent how pleasing or ugly the line is. The content of a representation must be what that representation contributes *in common* to the representations of which it is a part. And the content that a complex representation contributes to other representations of which it is a part is just what it inherits from the representations that are part of it.[2] Another way to put the point is that compositional representations have their contents "intrinsically," in the following sense: fixing the contents of the primitive constituents and their mode of combination thereby fixes the content of the complex representation they compose, in a strictly "context-independent" manner—that is, with no regard to what else is happening in the world or how the representation is related to any representations that are not its constituents (Fodor and Pylyshyn, 1988, Szabó, 2012).

A semantic theory of how perceptual representations are systematically related would be integral to a "meta-representational" account of why perceptual states have the contents that they do—what Neander (2017) calls the "content-determinacy" question. An answer to the content-determinacy question must explain both why the primitives in one's "perceptual dictionary" represent certain features rather than others and why their combinations have the semantic import that they do. Compositional theories do not answer these why-questions, but they place constraints on the acceptable answers. Any account of how the contents of the primitives are fixed must, for example, accommodate the contribution that those primitives make to more complex representations. Moreover, a compositional theory provides a framework for generalizing from an account of why one representation has the content that it does to an account of why systematically related representations have the contents that they do (Rescorla, 2019).

Semantic compositionality is a property of *representations*; it constrains, but does not determine the sorts of *processes* or "inferences" that form and operate upon those representations. Compositionality does not require, for example, that primitive representations are formed first and then combined at a later stage. One's representation of the scene might be computed by synthesizing prior representations of each line segment. Or one's visual system might first form a generic representation of a scene containing three line segments of different orientations, and then proceed to work out their specific orientations. Some combination of "analysis" and "synthesis" might take

[2] This explanatory role of content was a guiding consideration in Frege's conclusions concerning the denotations of number terms, names, predicates, and whole sentences, for example (Hintikka, 1983, Burge, 2005).

place in the course of arriving at a stable representation of the scene. Of course, certain sorts of representational processes are more apt for exploiting compositional representations than others. If a representation of three line segments is composed of three representations of individual line segments, then the former representation makes distinct contents about those individual line segments readily available for further processing. Visual search processes can query the representation for a constituent representation of a horizontal line segment on the left, for example.

Relatedly, the way perceptual representations epistemically support perceptual beliefs depends on how the contents of perception are taken up in belief, which in turn depends on how the structure of a perceptual state makes that content available for uptake. If my representation of the three line segments is composed from constituent representations of each line segment, then it is in principle possible to distinguish between the warrant that my perceptual state gives for the belief that the line segment on the left is horizontal and the warrant it gives for the belief that the line segment on the right is tilted. Perceptual beliefs may be warranted to different degrees by different parts of my perceptual state. If my perceptual state had a different structure, or if that structure had a different semantic import, or if the perceptual state were primitive (lacking constituents), then the explanation of how the perceptual warrant given by that state factors out would have to be different.

2.2 A Basic Theory

I will now sketch a basic semantics from which the analyses of some rudimentary percepts can be derived. I will motivate various refinements along the way, attending more to the shape of the theory than the formal details. Two features of this basic account are that constituent representations of the different aspects of a scene have the same type of semantic value (*semantic uniformity*) and these constituents can combine without substantive restrictions (*unrestricted combination*). Many semantic accounts of perception share these two features, to various approximations (for example, Clark, 2004, Fodor, 2008, Kulvicki, 2015). In the next sections, I will argue that we must reconsider both these features in order to account for core perceptual phenomena.

For now, let us continue to consider perceptual representations of stimuli that consist of three line segments arranged in a row, each of the same length and color. The segments can occupy one or the other of three positions: "position 1" on the left, "position 2" at the center, and "position 3" on the right. The segments can have one of 4 possible orientations: $0°$ (vertical), $45°$ (tilted right), $90°$ (horizontal), or $135°$ (tilted left). Reproducing the first example from above:

— | /

Viewing this stimulus under normal conditions, I visually represent a horizontal line (call it a) on the left in position 1, a vertical line (call it b) at the center in position 2, and a tilted line (call it c) at the right in position 3.[3] Call my current percept of this scene "$\mathbf{S_{abc}}$" (I use boldface characters to designate mental representations). A semantics of perception aims to specify the condition that must be satisfied for a perceptual representation to be veridical. For the purposes of distinguishing $\mathbf{S_{abc}}$ from other percepts of the stimulus set, we can take its veridicality condition to be:

> $\mathbf{S_{abc}}$ is perfectly accurate if and only if a is at position 1 and oriented $90°$ and b is at position 2 and oriented $0°$ and c is at position 3 and oriented $45°$.

This characterization of the veridicality condition of my perceptual state rests on several assumptions. First, I assume that perceptual representations function to be about particular property-bearing items in a scene (in this case, the line segments a, b, and c). Perceptual representations do not have existentially general contents (Burge, 2010, Schellenberg, 2018), nor do they merely place features at locations (positions 1, 2, or 3) without attributing them to particular entities (a, b, c) that occupy those locations (Clark, 2004, Matthen, 2004). Second, I am concerned with the *accuracy conditions*, rather than truth conditions, of perceptual states. Whereas truth is an all or nothing property of propositions, accuracy can come in degrees. If one misrepresents a as tilted $45°$, one would still be representing it more accurately than if one had represented it as tilted $0°$. Though I will only discuss the conditions on perfect accuracy here, models in perceptual psychology purport to explain the degrees to which perceptual states are accurate under different

[3] I assume that mature neurotypical humans and phylogenetically close primates will normally represent this stimulus in approximately the same way. Still, it is an idealization to talk of "a semantic theory of perception." Different individuals, of different species, at different times in their lives will represent things differently and so will require different semantic theories to capture their representational capacities. Moreover, there is unlikely to be one semantic theory that covers all perceptual capacities, or even all visual capacities.

conditions. Moreover, pictures and sub-propositional representations such as demonstrative noun phrases ("that round jar") can be evaluated for their accuracy even if they do not in themselves have truth conditions. And even where representations have both truth conditions and accuracy conditions, accuracy may be independent of truth. Suppose that in uttering "That round jar is empty," I thereby demonstratively refer to a square jar that I mistook as round. If the jar is empty, then by some accounts what I said is true even though it inaccurately characterizes the jar as round. Throughout, I remain neutral about whether perceptual states have propositional truth conditions in addition to accuracy conditions.

The goal is to specify the accuracy conditions for every perceptual representation one might have of the 64 possible stimulus types in this simple stimulus set. One approach would be to enumerate the accuracy conditions of each perceptual state. However, this approach is not only cumbersome, it fails to capture systematic patterns in how different stimulus configurations are represented. For example, each of the line segments in this stimulus set could in principle have been represented in a given way even if the other two line segments were represented differently. In order to represent a as at position 1 and horizontal, I need not represent b and c the way I do, or even at all. Psychological theories ascribe constituent structure to our percepts partly in order to explain how the way one aspect of the stimulus is represented does or does not depend on the way another aspect of the stimulus is represented (Lande, 2021). Nearly all psychological models would treat $\mathbf{S_{abc}}$ as a complex perceptual representation consisting of constituents, $\mathbf{c_a}$, $\mathbf{c_b}$, and $\mathbf{c_c}$, each representing a different distal contour or line segment.

Instead of enumerating the accuracy conditions of each percept, one can specify principles according to which representations of a given structural form derive their contents from the contents of their constituents. Suppose that $\mathbf{S_{abc}}$ is a "concatenation" of atomic representations, $\mathbf{c_a}+\mathbf{c_b}+\mathbf{c_c}$, where a perceptual concatenation of representations is accurate if and only if its constituents are accurate (see Sober, 1976).[4] To derive the contents of each total perceptual representation, the theory also needs to specify what "atomic" representations of line segments are possible and what their

[4]While concatenation clearly has the flavor of conjunction, it should not be assumed to be a logical operator. For example, it operates on degrees of accuracy, not on truth values.

accuracy conditions are. Here is a first pass at a theory, assuming that the atomic representations attribute both orientation and position to the items that they represent.

Concatenation: (1) For any atomic representations $\mathbf{c_i}$, $\mathbf{c_j}$, $\mathbf{c_k}$, $\ulcorner \mathbf{c_i} + \mathbf{c_j} + \mathbf{c_k} \urcorner$ is a structurally possible representation.

(2) $\ulcorner \mathbf{c_i} + \mathbf{c_j} + \mathbf{c_k} \urcorner$ is perfectly accurate iff $\mathbf{c_i}$ is perfectly accurate and $\mathbf{c_j}$ is perfectly accurate and $\mathbf{c_k}$ is perfectly accurate.

Atomic representations: (3) The atomic perceptual representations are $\mathbf{c_a}, \ldots, \mathbf{c_k}$.

(4) $\mathbf{c_a}$ is perfectly accurate iff a is at position 1 and oriented $90°$.

(5) $\mathbf{c_b}$ is perfectly accurate iff b is at position 2 and oriented $0°$.

(6) $\mathbf{c_c}$ is perfectly accurate iff c is at position 3 and oriented $45°$.

$\vdots$

Since I am assuming that perceptual representations function to be accurate of the particular items one encounters in a scene, there may be as many clauses for atomic representations as there are episodes in which one seems to encounter a distinct line segment. However, there is a sense in which different representations of particular line segments that are horizontal and at position 1 are of the same psychological kind. Different representations of horizontal lines at position 1 are the products of the same general capacity to single out line segments in that position and orientation, deployed on particular occasions to represent particular individuals in one's immediate environment. Though these representations may be of different particulars, they attribute a common set of features to those particulars (Burge, 2010). We would like to give a semantic characterization of the common kind (call it "$\mathbf{c}_\tau$") under which fall these specific representations (call them "$\mathbf{c}_{\mathbf{i}\tau}$"), while respecting the fact that each functions to be accurate of different particulars. One way of doing so is to provide *conditionalized* accuracy conditions (cf. Larson and Segal, 1995). For example:

If $\mathbf{c}_{\mathbf{i}\tau}$ is a representation of x_i, then $\mathbf{c}_{\mathbf{i}\tau}$ is perfectly accurate iff x_i is at position 1 and oriented $90°$.

I will set aside the question of what it takes for a perceptual state to be a representation of one particular item rather than another, except to say that it is plausibly a necessary condition that the item be causally responsible for the formation of the perceptual state (Strawson, 1979).

Can these atomic representations be analyzed even further? As I noted before, the fact that one can represent a horizontal line segment at any of the three positions is consistent with the hypothesis that we have a number of different primitives that represent specific pairs of orientations and locations. However, this hypothesis does not fit the evidence that capacities to represent orientation have psychological signatures that can differ from those of capacities to represent position. For example, the precision with which one represents a line segment's orientation may not be the same as the precision with which one represents its position. This difference between orientation and location perception can manifest in "illusory conjunctions," in which one misrepresents which orientations are located at which positions in the display (Treisman and Schmidt, 1982). For example, if I were distracted or had very little time to process the abc display above, I might have formed a different percept, $\mathbf{S_{bac}}$, as of a vertical line in position 1, a horizontal line in position 2, and a right-tilted line in position 3 (Treisman and Paterson, 1984, Golomb et al., 2014). Crucially, the finding is that I am liable sometimes to represent *the very same orientations*, but in different positions. Illusory conjunctions do not misrepresent stimuli wholesale. Rather, I can sometimes be more precise in my perception of orientation while independently being less precise in my perception of position, with the consequence that sometimes I get the orientations right but swap their positions in the display. If orientation and position were coded by the same primitive variable in my visual system, one would expect orientation and position to be perceived necessarily with the same variance (Lande, 2021). A good explanation of how the precision with which one perceives an item's orientation can vary separately from the precision with which one perceives that item's position is that these features are each coded by separate constituents with their own psychological traits (Treisman, 1986, Matthen, 2005).

Accordingly, suppose that atomic representations of contour segments are themselves complex, consisting of primitive representations $\mathbf{c_{i\rho}}$ of an item's position and $\mathbf{c_{i\theta}}$ of its orientation. For example, $\mathbf{c_a}$ has as constituents $\mathbf{c_{a1}}$, representing a at position 1, and $\mathbf{c_{a90^\circ}}$, representing a as horizontal. The claim is not that one can represent orientation *apart* from position. It may be that one can only ever represent orientation at *some* location or other. So, distinguish between *atomic* perceptual representations and their sub-atomic primitives. Atomic representations are the least complex representations that can occur without being combined with other representations—they are the simplest "well-formed" representations in a perceptual system. Sub-atomic primitives are the least complex representations that can vary apart from each other, but which cannot occur except in combination with other representations . We can model the atomic representations of line segments as 2-dimensional vectors, $\ulcorner\langle \mathbf{c_{i\rho}}, \mathbf{c_{i\theta}} \rangle\urcorner$. While these can be decomposed into independent components, the first component must always occur in the context of a second component.

As a compositional rule, when an orientation representation is bound to a position representation, that combination is accurate just in case both of the primitives are accurate of the same item—just in case that same item has that orientation and that position. Combining primitive representations has the semantic import that those representations are of the same item. In fact, illusory conjunctions demonstrate that a representation of an element might be composed from a primitive representation $\mathbf{c_{i\theta}}$ of that element's orientation together with a primitive representation $\mathbf{c_{j\rho}}$ of some other element's position. The primitives that make up an atomic representation do not have to be "co-indexed," so to speak. 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. But another way for it to be inaccurate is for it to be a combination of an accurate representation of one item's orientation and an accurate representation of a different item's position.

Finally, consider the semantics of the primitive representations themselves. While compositional principles determine the contents of complex representations, "analog" relations between primitives can play a role in determining the contents of the primitives. Psychological models typically treat primitive perceptual representations as related along dimensions of psychological similarity that correspond to relations or dimensions of represented items in the world (Gauker, 2012, Beck, 2019). For example, just as position 1 is closer to position 2 than to position 3, the *representation* $\mathbf{c_{i1}}$ might be more psychologically similar to the representation $\mathbf{c_{i2}}$ than it is to $\mathbf{c_{i3}}$. Let μ be a systematic mapping from the positions of primitive representations in psychological similarity space to the positions of distal features of the world on explanatorily significant dimensions of objective similarity, such that relations of psychological similarity or "distance" between primitives corresponds to the objective similarity in represented features. Then instead of separately specifying the accuracy conditions of each positional representation, one can just specify that $\mathbf{c_{i\rho}}$ is perfectly accurate just in case the item that it represents is located at position $\mu(\rho)$, and likewise for primitive representations of orientation. (Throughout the rest of the paper, I will make the idealization that μ is the identity function.)

Here, then, is the basic theory in full:

Primitives: (1) For all $i \in I$ (where I is a set of indices, such as a, b, c, for indicating specific perceptual states), $\rho \in \{1, 2, 3\}, \theta \in \{0, 45, 90, 135\}$, $\mathbf{c_{i\rho}}$ and $\mathbf{c_{i\theta}}$ are primitive representations.

(2) If $\mathbf{c_{i\rho}}$ is a representation of x_i, then it is perfectly accurate iff x_i is located at position $\mu(\rho)$.

(3) If $\mathbf{c_{i\theta}}$ is a representation of x_i, then it is perfectly accurate iff x_i has orientation $\mu(\theta)^\circ$.

Feature Integration: (4) For any $\mathbf{c_{i\rho}}$ and $\mathbf{c_{j\theta}}$, $\ulcorner\langle\mathbf{c_{i\rho}}, \mathbf{c_{j\theta}}\rangle\urcorner$ is a structurally possible representation.

(5) If $\ulcorner\langle\mathbf{c_{i\rho}}, \mathbf{c_{j\theta}}\rangle\urcorner$ is a representation of x_k, then it is perfectly accurate iff (a) $\mathbf{c_{i\rho}}$ is perfectly accurate and $\mathbf{c_{j\theta}}$ is perfectly accurate, and (b) if $\mathbf{c_{i\rho}}$ is a representation of x_i and $\mathbf{c_{j\theta}}$ is a representation of x_j, then $x_i = x_j = x_k$.

Concatenation: (6) For any structurally possible representations α, β, and γ, $\ulcorner\alpha + \beta + \gamma\urcorner$ is a structurally possible representation.

(7) $\ulcorner\alpha + \beta + \gamma\urcorner$ is perfectly accurate iff α is perfectly accurate and β is perfectly accurate and γ is perfectly accurate.

It is trivial to extend the theory to accommodate representations of indefinitely large, two- or three-dimensional displays of line segments with indefinitely fine-grained orientations and positions. To accommodate representations of items with various colors, sizes, shapes (blobs, corners, junctions), suppose that additional types of primitives are included in the visual dictionary and can integrate into the atomic representations. Note that the theory as it stands permits the representation of multiple line segments located at the same location (for example, "+" and "×"), though it does not account for the ability to represent these segments as parts of the same cohesive item.

Recall the two features of the theory that I mentioned at the outset. First, the theory characterizes the constituents of percepts as *semantically uniform* in that every constituent has the same type of semantic value. Just as the representations in propositional logic all have truth values, according to this theory all the constituents of a perceptual representation all function to have accuracy values (given a specification of the items that they represent). Each constituent of a perceptual representation determines a complete condition on what the represented item would have to be like for the constituent to be perfectly accurate. This is true even of the sub-atomic primitives. While the representation of an item's orientation cannot, according to the theory, occur except in combination with a representation of that item's location, nevertheless these representations can each in themselves be evaluated for their accuracy (cf. Shea, 2018, p. 163–4). Second, these constituents can in principle *combine freely*, subject to minimal restrictions on type. As the theory stands, there are no substantive restrictions on which location primitives can bind with which orientation primitives or on which atoms can concatenate with each other.

One substantial point of debate concerns whether there are separate primitives for singular and attributive elements of perceptual representations and, if so, whether these enter into predicational structures or not.[5] The present theory assumes that perceptual representations ultimately function to be accurate of particular objects and so can have singular accuracy conditions. However, the theory so far is agnostic about whether representations of position and orientation ($\mathbf{c}_{\mathbf{i}\rho}$ and $\mathbf{c}_{\mathbf{j}\theta}$) are composed from more basic singular elements ($\mathbf{i}$ and $\mathbf{j}$, say) and distinct general elements ($\mathbf{c}$,

[5]For some recent discussions, see Clark, 2004, Fodor, 2008, Burge, 2010, Matthen, 2014, Schellenberg, 2018.

ρ and θ). If the theory is extended to incorporate these distinct types of elements as primitives, then semantic uniformity would not extend to these basic elements, though it would remain the case that any *complex* constituent, representing one aspect of a scene by combining a singular element and general elements, would have the same semantic role as any other complex constituent that represents any other aspect of the scene. (By contrast, in English for example it is not just primitive nouns and verbs that can have different types of semantic values, but more complex phrases too.) Moreover, if there are distinct singular and general constituents, the theory is agnostic about whether these are combined *predicatively* so as to yield propositional truth conditions in addition to the accuracy conditions specified by the theory.

I want briefly to indicate how this semantic theory relates to other types of projects that are concerned with the nature of perceptual representation. One project is to say what sorts of abstract objects perceptual contents are. For example, Peacocke (1992) identifies perceptual contents with sets of what he calls "scenes." Others have argued for identifying perceptual contents with sets of possible worlds (Stalnaker, 1998). These sorts of views say little about how such contents are compositionally derived according to the way the representational vehicle itself is structured. By contrast, the primary aim of the theory sketched here is not to give a metaphysics of perceptual contents, but rather to characterize the systematic relationships between such contents (whatever they turn out to be). The theory describes these systematic relationships by characterizing how structural relationships between representations induce semantic operations on accuracy conditions.

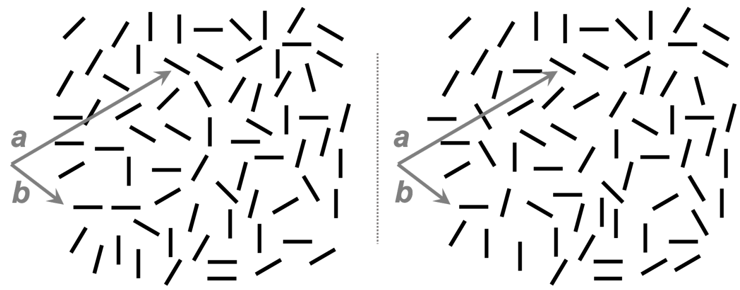
Another project more concerned with the compositional apparatus of perception is dedicated to settling whether this apparatus is analog and/or iconic or whether it is symbolic and/or sentence-like in nature. Many take the answers to these questions to depend on whether there are distinct singular and general elements in perception that combine through predication. As I have said, the current theory is agnostic on that point. In any case, it remains a contentious subject exactly how to carve up distinctions like the one between iconic and symbolic representations. I will remain neutral about where the current semantic theory falls or how the properties of semantic uniformity and unrestricted combination bear on this issue. The project in what follows is to see how the theory and others like it have to be revised in order to analyze a fuller range of perceptual phenomena. This project can be pursued quite far—the relevant puzzles and results can be stated in some detail—without settling the more foundational questions raised in these last few paragraphs.

3 Emergent Content and Context

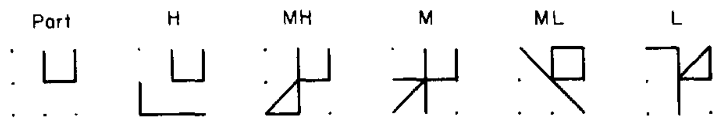
The theory sketched in the previous section implies that perceptual representations of different aspects of a scene, and the combinations of these representations, are in large part semantically uniform and freely combinable. The theory does not stray far from the views that Wertheimer and his colleagues rebuffed, according to which percepts are mere "and-summations" of their parts. But the explanatory power and generality of the theory and others like it are at best unclear, for reasons that were center stage in the Gestalt school. The theory, as it stands, does not account for the representation of "emergent," "global," or higher-order features of items or for the context effects involved in representing such features.

To a first approximation, emergent features are features that are perceptually attributed to collections of represented items or features in a scene, or which are attributed to individual items in virtue of their relationship to other represented items in a collection (Pomerantz and Cragin, 2015). These features include the feature of being an extended contour, forming a closed boundary, being symmetrical about an axis, having an "average" color or orientation, or even being an "odd one out" (see Palmer, 1977, Brady et al., 2011). For example, one's representation of left panel of Figure 1a goes beyond concatenating representations of the locations and orientations of different line segments; one represents a specific subset of line segments, including a and b, as forming a common *contour*. One's perceptual representation of the left panel is accurate only if there is a contour of which both a and b are parts. But the basic theory, as it stands, does not yield this accuracy condition.

One could posit that the visual dictionary contains primitive representations of various whole contours, which can be concatenated with one's total percept of the scene. If the representation of the whole contour were merely an additional primitive concatenated into the percept, one would



(a)



(b)

Figure 1: Emergent features and context effects in perceptual organization: (a) The two line segments, a and b, in the left panel appear to be part of a common contour, while a physically identical pair of line segments in the right panel does not (based on Geisler and Super, 2000, p. 682). (b) The "U"-shaped "Part" can be perceived as a discrete cohesive part of panels H, MH, *and* M, *but not* ML *or* L *(from Palmer, 1977, p. 452).*

expect it to be possible, in principle, to token the representation of the whole contour without tokening the representations of the segments. However, moving or removing the segments dramatically interferes with one's ability to represent the whole contour (Field et al., 1993). Moreover, even when the segments are represented as part of a common whole, representations of the individual line segments are separable—they can have separate psychological signatures, including different levels of precision or noise (Pomerantz and Pristach, 1989, Brady and Alvarez, 2015). For example, under certain circumstances when viewing the display I may misrepresent one segment or another as in a different location, even belonging to another object (Treisman and Paterson, 1984).

The hypothesis that the representations of a and b are distinct constituents of the representation of the whole contour would explain both why the representation of the whole contour cannot occur without the constituent representations of a and b and why the constituent representations of a and b could nevertheless have separable signatures (Lande, 2021). One way to frame this hypothesis is to suppose that there is a type of representation of the higher-order attribute of being a contour that *must* be combined with representations of line segments so as to represent those line segments as parts of a common contour. Alternatively, representations of extended contours might be thought of as composed from representations of individual segments according to a distinctive mode of combination—a contour-specific integration operator. Either way, the basic semantic theory will have to be expanded to characterize how representations of higher-order attributes combine with or are composed from these more basic representations.

The representation of emergent features is closely related to the ubiquity of perceptual context effects. Perceptual context effects arise when the way one aspect of a scene is represented depends on the way other aspects of the scene are represented (see Schwartz et al., 2007, Todorović, 2010). It is not the case that the same atomic representations of line segments with specific orientations and positions can compose the same perceptual content no matter the context in which these representations are embedded. One represents a and b as belonging to a common contour in the context of representing the "connecting" line segments in the left panel of Figure 1a. When one does not represent connecting segments, as in the right panel, one does not represent a and b as part of a common contour. As Wertheimer put it, the connecting lines are "pro-structural" elements that enable one to represent the segments as part of a common contour. By contrast, "contra-structural" elements preclude the perception of a whole contour. While one can distinctly represent a coherent U-shaped contour in panels "H," "MH," and "M" of Figure 1b, it is extremely difficult to discern this same contour as a cohesive part in panels "ML" and "L," though one can individually discriminate the very same component line segments (Palmer, 1977, Ankrum and Palmer, 1991). The representations of those segments, $\mathbf{c_i}, \mathbf{c_j}, \mathbf{c_k}$, can compose a representation of a contour (or can combine in some way with a representation of a contour attribute) in some contexts but not others. More generally, one cannot attribute higher-order features to arbitrary sets of items, and the higher-order features that one attributes to some items depends on how one represents other items in the scene.

Context effects also arise when representations of first-order features of items are influenced by one's representations of higher-order features. For example, "iso-feature suppression" occurs when representations of a given type of feature make it harder to represent the same or similar features

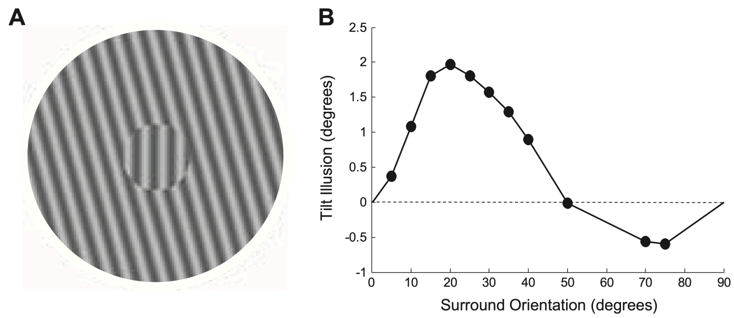


Figure 2: The simultaneous tilt effect for gratings of lines. (a) Though the central grating is physically vertical, it appears to be tilted slightly clockwise. (b) When the center and surround have physically similar orientations, their perceived orientations are "repulsed"; when they have substantially different orientations, their perceived orientations can be perceptually "attracted" (from Clifford, 2014, p. 4).

in the scene (Li, 1999). It is more difficult to represent a line segment as having a given orientation if it is surrounded by line segments of similar orientations than if it is presented in isolation. As a result, contrasting features are often emphasized in perception. It turns out that these effects are especially operative when one represents the features as belonging to the same object or group. For example, in the "simultaneous tilt effect," the perceived orientation of a contour or grating is "repulsed" from the perceived orientation of its surround when these differ slightly (Figure 2). The strength of this effect is a function of the perceptual evidence that the target and the surround belong to the same object: the effect is substantially weaker if one does not perceive the target and the surround as part of the same surface or object. A higher level type of context effect occurs when the perceived orientation, position, or motion of an item depends on the representation of another item or configuration that serves as a "reference frame." For example, in Figure 3, one tends to see the triangle as oriented to the right if embedded in the horizontal rectangle or as oriented to the upper left if embedded in the tilted rectangle.

"Context," as I am using it here, does not refer to the distal environment in which a representation is tokened (its *context of tokening*; analogous to the "context of utterance" in the case of language), but rather refers to the type of representation with which some constituent is combined (the constituent's *representational context* or *context of embedding*; analogous to sentential and discourse contexts in language). Contrary to the basic theory offered above, it is not the case that a

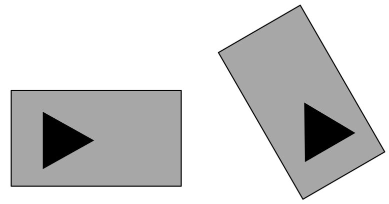


Figure 3: Reference frame effects: an equilateral triangle appears to be pointing to the right when enclosed in a horizontal rectangle, while a physically identical triangle appears to point to the upper left when enclosed in a tilted rectangle.

given perceptual content can freely occur in arbitrarily different contexts of embedding. As Camp suggested, perceptual abilities to represent contours, orientations, and other features are "highly local and context-dependent."

How can we accommodate these paradigmatically "holistic" phenomena within a compositional framework? In particular, if the content of a representation depends only on the contents of its own constituents and how they are related, how can the representation of an item depend on that representation's context of embedding? Broadly, there are three ways to give a compositional analysis of context effects. First, the same set of constituents might stand in different structural relationships, and so form a different type of construction, depending on the context in which they are embedded. I will argue that the same representations of contour segments can combine into a representation of a whole contour in some contexts but not others. An analogy here is to structural disambiguation in language: the words that make up the string, "Kiki and Jiji," are part of a common constituent in the context of the sentence "Tombo saw Kiki and Jiji," while the same words are parts of distinct constituents in the sentence "Tombo saw Kiki and Jiji jumped."

Second, representational context might bias which representation of a stimulus gets formed in the first place. I will argue that the tilt effect arises when a representation of an oriented surround "selects for" contrasting representations of the center's orientation. Here, an analogy can be made with the role that sentential context can play in lexical disambiguation: the word-form "duck" is mapped to one lexical item in the sentence "Kiki saw a duck" while it is mapped to a different lexical item in the sentence "Kiki had to duck." The basic theory sketched above makes no provisions for anything like structural or lexical disambiguation. Either type of analysis for perceptual context effects requires giving up the presumption that perceptual representations can combine without substantive restriction.

Third, context effects might arise in virtue of the different semantic roles that perceptual representations can have. I will argue that reference frame effects arise when one perceptually represents something as having a specific orientation *relative to y*, where the content of this representation functions to be integrated with the content of a representation that supplies a reference item, y. Likewise in language, adjectival modifiers such as "former" function to combine with common nouns such as "champion" or "professor" so as to represent someone as no longer a member of whatever class is denoted by that noun. This sort of analysis requires distinguishing between different types of semantic roles that perceptual representations can have. Just as "former" cannot in itself be an accurate or inaccurate characterization of an individual, so too a representation of an item's *relative orientation* does not in itself determine a complete accuracy condition. The content of this representation constitutively functions to integrate with the content of other representations.

In the next section, I will motivate and elaborate these analyses of perceptual context effects. In Section 5, I will argue that these particular analyses all reveal a close relationship between context effects and the objectivity of perception. The discussion here broadly echoes one in natural language semantics. Semantic theories of language typically start by covering relatively simple fragments of a language. These analyses usually require refinements and revisions in order to deal with context effects. For example, the meanings of quantifier phrases, adjectival constructions, conditionals, tense clauses, and anaphoric pronouns appear to shift depending on the broader sentential and discourse contexts in which these constructions are embedded. If the content of a construction is derived compositionally from the contents of its parts and their arrangement, then that construction's meaning must be independent of its relation to other external constructions. So, if a phrase appears to take different meanings when embedded in different sentences, then a compositional analysis would have to show how superficially similar surface strings differ in their underlying syntax when they occur in different sentences, in line with familiar examples of syntactic and lexical disambiguation, or else how, despite appearances, these strings in fact have the same meaning that functions to integrate with the meaning of the phrase with which it is combined (see Partee, 2004, Janssen, 2011). Some doubt that such analyses will always be plausible. Hintikka, for example, writes 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" (Hintikka, 1983, p. 265; see also Higginbotham, 1986, 2003). Camp's skepticism about the prospects of semantic theories of pictures and perceptual representations echo these doubts about the compositionality of context effects in language. By contrast, my approach in what follows is more aligned with those who argue that sophisticated compositional analyses along the lines indicated above are in fact central to understanding why context effects arise (for example Partee, 2007).

4 The Semantics of Context Effects

I now sketch compositional explanations of the three context effects introduced in the previous section: context effects in contour perception, tilt effects, and reference frame effects. These effects have been studied extensively and each provides an illuminating example of a way to analyze context effects within a compositional framework. In particular, each illustrates how an adequate compositional semantics of perception must give up commitments either to unrestricted combination or semantic uniformity.

4.1 Contour Integration

Consider one's perceptual representations of the left and right panels in Figure 1a. We want to account for the fact that one's percept of the left panel is accurate only if a and b belong to the same contour while one's percept of the right panel can be accurate even if a and b do not belong to the same contour. There is no reason, on the minimal theory, to expect the atomic representations of a and b to be different between the two panels. Moreover, there is no way, on this theory, to distinguish between the content of $\mathbf{c_a} + \mathbf{c_b}$ as it occurs in the representation of one panel and the content of $\mathbf{c_a} + \mathbf{c_b}$ as it occurs in the representation of the other. Of course, one represents the rest of the display quite differently. We can distinguish between the accuracy conditions of the total representation $\mathbf{c_a} + \mathbf{c_b} + \mathbf{\Gamma}$ of the left panel and of the total representation $\mathbf{c_a} + \mathbf{c_b} + \mathbf{\Gamma}'$

of the right panel, where $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ comprise the representations of all the remaining items in the respective panels. However, this difference still will not encompass that $\mathbf{c_a} + \mathbf{c_b} + \mathbf{\Gamma}$, unlike $\mathbf{c_a} + \mathbf{c_b} + \mathbf{\Gamma}'$, is accurate only if a and b are part of the same contour. Indeed, ambiguous stimuli such as a matrix of dots can give rise to different percepts of grouping, into columns or into rows, without any difference in the basic items that one represents. To explain the context effect requires ascribing additional structure to the perceptual representations.

One way to amend the theory is to suppose that there is a primitive representation of a higher-order attribute, call it "$\mathbf{CONTOUR}$," that must be combined with representations of individual contour segments so as to represent a whole contour, x, to which those segments belong. Such an amendment would require specifying not just the content of this representation, but also how it combines with representations of individual segments. This primitive cannot just be concatenated with the representations of the individual items. Such concatenation would fail to capture the fact that the representation of the whole contour *requires* the representation of the segments—the representation of the whole contour is not self-standing and is not just tacked on alongside the rest of the constituent representations of segments. Moreover, we need some way to capture the fact that a, b, and certain other line segments are represented as part of the contour, while many others are not. The integration and concatenation operations defined above are not fit for the job. There must be a distinctive way of selectively combining certain representations of individual items to represent the higher-order feature of being a contour. To simplify things for present purposes, I will eschew talk of a $\mathbf{CONTOUR}$ primitive and the attendant questions that arise about how this primitive combines with representations of contour segments. Instead, I will pursue the idea that there is a contour-specific mode of combination, a contour-construction or *contour operator*, $\ulcorner \mathbf{c_i} \circ \ldots \circ \mathbf{c_n} \urcorner$, by which representations $\mathbf{c_i}, \ldots, \mathbf{c_n}$ of individual contour segments compose representations of extended contours.[6] Whereas the representations of a and b enter into such a structural relationship

[6]This approach is analogous to a semantics that treats a term "syncategorematically" as a grammatical particle rather than "categorematically" as an item in the lexicon (see Quine, 1986, p. 26–30). Both approaches can yield the same veridicality conditions. I will not discuss the explanatory considerations that would favor one approach over the other. For now, I pursue the "contour operator" approach for simplicity. Whether through construction with a contour attributive or through a contour-specific mode of combination, representations of contour segments must be able to relate in a different way than through the integration and concatenation operations defined above.

in one's percept of the left panel, they do not in one's percept of the right panel. An adequate analysis of the context effect should provide evidence for such a structural relation and explain why constituents can stand in that relation in some contexts but not others.

4.1.1 Evidence for Contour Structure

The critical evidence that the percepts of a and b are structurally related in the left panel in a way they are not in the right panel comes from studies of *grouping advantages*. Normally one is substantially more accurate, more precise, and faster at detecting, identifying, attending to, and holding in memory sets of items or properties that are perceived as unified—as parts or features of a common entity or group—compared with sets of items or properties that are not perceived as unified. Sets of contour segments that are perceived as belonging to the same contour are more visually salient and more quickly, accurately, and precisely detected than sets of contour segments that are not perceived as unified (Field et al., 1993). One is also more quick, accurate, and precise at detecting segments when they are perceived as part of a unified whole contour (Kapadia et al., 1995), at shifting attention between parts of a common whole (Barenholtz and Feldman, 2003, Kimchi et al., 2015), and at discriminating between features of those parts, such as their orientations (Kempgens et al., 2013). Similar advantages accrue to memorizing and recalling features of elements that one perceives as parts of a common whole. Grouping advantages 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 representation of unity corresponds to the unity of representations. Individual elements are visually represented as parts of a unified whole if and only if the representations of those elements are constituents of a common representation. It is easier and more efficient to encode, store, and manipulate sets of representations that are structurally related than those that are structurally unrelated. 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. So, grouping advantages obtain for items that are perceived as belonging to a common whole because the representations of those items are in construction with each other.

Whether a set of elements is perceived as unified, therefore giving rise to grouping advantages, depends on how one represents other elements in the scene. Palmer (1977) asked subjects to report whether three-item sets of line segments were present or not in larger six-item sets of line segments. He found that subjects are categorically faster and more accurate in detecting the presence of a subset of line segments, such as the line segments making up the "U"-shaped part in Figure 1b, when one sees those segments as forming a coherent, complete part of a figure (as in panels *H*, *MH*, and *M*) than in contexts where one sees those segments as belonging to different parts (for example, *ML* and *L*; see also Palmer and Beck, 2007, Ankrum and Palmer, 1991, Brady et al., 2011). Grouping advantages do not depend directly on the other line segments one perceives in the scene. Ambiguous stimuli, such as arrays of dots that can be grouped either as columns or as rows, can give rise to different percepts of unity and different corresponding grouping advantages, despite representing all the same basic items. When the perceived unity, and corresponding grouping advantages, of a set of items varies from context to context, it is because the different contexts influence the representation of the target items: only in some contexts do the representations of those items form a representation of a whole contour.

4.1.2 Compositional Analysis

Why are the same representations integrated into a contour representation in some contexts but not others? To answer this, we can turn to computational models of contour integration. Despite the artificiality of Figure 1, contour integration is a critical perceptual achievement. The bounding contours of objects often are interrupted by occluding surfaces, shadows, highlights, or matching textures, and yet we effortlessly distinguish the unified boundaries of these different objects. Even when the different surfaces in a scene have fully visible, uninterrupted boundaries, the light at the eye does not determine which edges belong to the same objects. Models of contour integration seek to explain how the visual system accurately represents which contour segments belong to the same object. These models standardly treat representations of the bounding contours along the outline of an object as recursively structured from more basic representations, $\mathbf{c_i}$, of minimal contour segments at a given spatial scale. Elder and Goldberg (2002), for example, model representations of extended contours as ordered sets of representations of contour segments: $\ulcorner(\alpha_1, \ldots, \alpha_n)\urcorner$. Let each α_i be either a complex representation of an extended contour or else an atomic representation of a contour segment, $\mathbf{c_i}$. It is not the case that any arbitrary sequence of contour representations, of randomly scattered contour segments for example, can form a representation as of a cohesive contour. Models of contour integration specify constraints on how representations of contour segments can be combined.[7] We can formulate these constraints in terms of a contour operator, $\ulcorner\alpha_1 \circ \ldots \circ \alpha_n\urcorner$, which returns a complex representation, $\ulcorner(\alpha_1, \ldots, \alpha_n)\urcorner$, just in case the constituents satisfy the conditions of the operator.[8] These complex contour representations are accurate if and only if their constituents are accurate and represent parts of the contour represented by the whole:

Contour Integration: (4) If $\ulcorner\alpha_1 \circ \ldots \circ \alpha_n\urcorner$ is a representation of x, then it is perfectly accurate iff for each immediate constituent α_i ($1 \leqslant i \leqslant n$):

(a) α_i is perfectly accurate, and

(b) if α_i is a representation of y, then y is a part of x.

The explanation of the context effects rests on the constraints on the contour operator. Most models of contour integration incorporate some version of the classic Gestalt principle of "good continuation" as a constraint on how representations of contour segments can combine into a representation of an extended contour. Field et al. (1993)'s "association field" is a classic example. The association field embodies a set of rules or constraints for how a given contour representation can combine with another, as a function of how well-aligned the represented positions and orientations of those contours are.[9] The association field systematically constrains which complex contour representations are possible as a function of what its constituents are.

[7]The psychologist Brian Keane 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 integration]" (Keane, 2018, p. 5).

[8]Cf. Geisler and Super (2000)'s "grouping operator."

[9]As Hess et al. (2015) describes it, contour integration only occurs when the represented orientations of the edges differ by no more than 60° and their represented positions are no farther apart than 4-6 edge lengths (for closely related rules on when contours can be amodally completed, see Kellman and Shipley, 1991).

Contemporary probabilistic and Bayesian models of contour integration treat these constraints more softly. In these models, the contour operator or association field can be thought of as determining the "binding strengths" between contour representations, as a function of how well aligned the represented contours are (cf. Geisler et al., 2001, Geman et al., 2002, Zhu and Mumford, 2006). As a first pass, the binding strength of a set of representations characterizes the degree to which these representations can enter into construction together—what one might call their degree of well-formedness. Geisler et al. (2001) and Elder and Goldberg (2002) define binding strengths in terms of conditional probability distributions. To a first approximation, the binding strength of two edge representations, $\mathbf{c_1}$ and $\mathbf{c_2}$, is the ratio of the likelihood that these two representations could be the constituents of a common contour representation, $P(\alpha = \mathbf{c_1}, \beta = \mathbf{c_2} | \ulcorner \alpha \circ \beta \urcorner)$, to the likelihood that these two representations could occur as unrelated constituents of a total percept, $P(\alpha = \mathbf{c_1}, \beta = \mathbf{c_2} | \sim \ulcorner \alpha \circ \beta \urcorner)$. These "likelihoods" need not, in principle, correspond to either the probability that a given combination actually *will* occur, which depends on a number of factors including specific stimulus cues and prior expectations, or the probability that such a combination would be *accurate*. Like prior models, these probabilistic models of the association field systematically constrain the formability of contour representations, now as a matter of degree, strictly as a function of what their constituents are—namely, as a function of the alignment between the represented positions and orientations.

Following, Elder and Goldberg (2002), the binding strength of a complex contour representation can be factored into the binding strengths of each succeeding pair of constituents of the contour representation. The more aligned the positions and orientations represented by $\mathbf{c_1}$ and $\mathbf{c_2}$, the more these representations are able to combine. So, the degree to which a contour representation is structurally possible depends on the alignment between pairs of represented, neighboring segments along the contour.[10] At a high level of abstraction, the constraint on combining representations of contour segments into representations of whole contours is something like this:

Contour Integration: (5) (a) The binding strength of $\ulcorner \mathbf{c}_{\rho\theta} \circ \mathbf{c}_{\rho'\theta'} \urcorner$ is proportional to the align-

[10] As Elder et al. (2018) suggests, this is an idealization: the binding strength may depend also on whether, for example, the whole contour can be represented as closed.

ment between (ρ, θ) and (ρ', θ').

(b) The binding strength of $\ulcorner \alpha_1 \circ \ldots \circ \alpha_n \urcorner$ is a product of the binding strengths of each pair, α_i, α_{i+1} $(1 \leqslant i \leqslant n-1)$.

Disregarding the different ways that empirical theories mathematically define *alignment*, all models will take the representations of a and b in Figure 1a to have a binding strength approaching zero. That is, $\mathbf{c_a}$ and $\mathbf{c_b}$ are not, in themselves, combinable (or: they have a vanishingly small capacity to combine). However, one's representation of a in the left panel can combine with one's representation of the adjacent element to its lower right, which in turn can combine with one's representation of that element's neighbor, and so on. So, $\mathbf{c_a}$ and $\mathbf{c_b}$ are combinable in the context of combining also with representations of these connecting elements, $\mathbf{c_a} \circ \ldots \circ \mathbf{c_b}$. Since there is no clear path from a to b in the right panel, the representations of a and b do not come into construction with each other and one does not therefore represent them as parts of a common contour. In the case of Figure 1b, the addition of "contra-structural" elements can destroy the perceived unity of a set of elements because in these contexts other combinations of representations are much stronger or simpler (Palmer, 1977; see also Geisler and Super, 2000, Feldman et al., 2014).

Not all contour representations are combinable, and those that are combinable are not equally so. As I argue elsewhere (Lande, 2021), part of what marks the association field as a *structural* constraint—part of a "contour grammar"—is that it systematically constrains which complex representations of contours can psychologically occur as a function of what their constituents are. This constraint abstracts from the variable input and background conditions that drive the formation of a specific representation on a particular occasion. Of course, the association field is part of a broader model of how a particular representation is formed on the basis of a particular sensory input and background conditions—an account of "contour inference." Such an account will include other principles by which representations of contour segments are computed from specific sensory cues, as well as principles by which prior expectations, task goals, and attentional selection can influence the formation of one representation rather than another.[11] However, even abstracting from these

[11] In the Bayesian models described above, what I am describing as a generative grammar of contours is encoded in

aspects of processing, there is a substantive structural constraint on the combinability of contour representations. This constraint suffices to explain the context effect.

4.1.3 Representational Function of the Context Effect

The substantive constraints on how contour representations can combine serve a representational function in light of the semantic import that such combinations have. A principal discovery in contemporary research on perceptual organization is that the constraints on perceptual organization correspond closely to the statistical relationships between parts of a common object (Brunswik and Kamiya, 1953, Geisler et al., 2001, Elder and Goldberg, 2002). Edges that are nearby each other and aligned, or that are part of a sequence of pairs of such edges, generally are far more likely to be part of a common contour than not. Substantially misaligned edges that are far apart and not connected by any smoothly linking edges are very unlikely to be part of a common contour. The constraints on how representations of contour segments can combine in different contexts ensure that composite representations of whole contours will have some minimal reliability of being accurate in the perceiver's normal environment. The representations that are least able to combine in a given representational context are also the ones whose combination is least likely to be accurate in that kind of context. These combinatorial constraints therefore constitute one aspect of the visual system's pursuit of representational accuracy.

It is a contingent fact that our visual contour grammar is coordinated in this way with the actual statistics of contours. In principle, the representational combinability of contour representations need not have corresponded to the objective combinability of contours in the world. The well-formedness of an English sentence, for example, is not a reliable guide to its truth. If the present analysis is right, then the combinability of a set of contour representations *is* a guide to the plausibility of the representation that they compose. In terms of the Bayesian models described above, an optimal system for computing accurate representations of natural contours is one in which a a generative grammar for how contour representations can be composed from representations of component segments models how real-world contours are normally organized from component

the likelihoods, while these other factors are encoded as input data, priors, and cost functions.

segments (cf. Yuille and Kersten, 2006, Tenenbaum et al., 2011).

4.2 Simultaneous Tilt Effect

I argued that in the case above a given set of basic representations might be combinable into a representation of a higher-order feature (a contour) in one context of embedding but not in another. Context can also influence how one represents the first-order features of individual items. In the simultaneous tilt effect, for example, the perceived orientation of a contour or grating is "repulsed" from the perceived orientation of its surround when the physical orientations of these differ slightly, and is "attracted" when the orientations are nearer to orthogonal (Figure 2). In principle, there are a number of different ways one might analyze the tilt effect.

One possibility is that the tilt effect is a result of a semantic operation on the contents of constituent representations of the central and surround stimuli. Plausibly, perceptual representations can have indeterminate contents (see, for example, Nanay, 2020). Suppose the representation of the central contour or grating attributes an indeterminate orientation, or a range of orientations. The compositional result of combining this representation with representations of surrounding contours might be a complex representation that attributes more determinate orientations to the center and surround contours, such that these more determinate orientations exaggerate the difference between the center and surround. While this particular analysis is in principle available, it is empirically implausible. As the title of Solomon and Morgan (2009)'s paper, "Strong Tilt Illusions Always Reduce Orientation Acuity," indicates, one is in fact able to discriminate the orientation of the central stimulus much more finely when it is not subject to a strong tilt effect.

Instead of locating tilt effects in a semantic operation on contents, we should look to structural constraints on how representations of the center can best combine with representations of the surround. Representations and their combinations are formed under conditions of uncertainty. Given the inevitability of noise, the proximal stimulus arising from the central contour or grating will always permit a number of distinct orientation representations. In viewing Figure 2a, the visual system will register some sensory cues indicating that the central contours are vertical and some cues consistent with those contours being tilted. The system may also register that the central contours are part of the same pattern as the surrounding contours. The representations of the surround may be biased against combining with representations of the central contours as vertical and toward combining with representations of those contours as tilted away from the surround. As a result, these more contrastive representations are selected to combine with the representation of the surround. I will call this sort of "disambiguation" by context, *structural selection*.

4.2.1 Evidence for Selection

I will briefly survey the evidence (a) that the tilt effect involves the formation of distinct representations that attribute different orientations, and (b) that the tilt effect depends on the representation of one stimulus being combined with the representation of the other.

As early as the primary visual cortex (V1), different patterns of neural activity arise in populations that are tuned to the orientation of stimuli in one position of the visual field, depending on activity in populations that are tuned to the orientations of surrounding stimuli (Gilbert and Wiesel, 1990, Kapadia et al., 2000). Differences in the activity of a neural population do not in themselves imply differences in the representations that those patterns realize or cause. However, in this case the differences in neural activity have a clear perceptual significance. When the visual system registers a vertical stimulus, a host of neurons will fire, with the strongest response in neurons that are tuned to vertical orientations and gradually weaker responses in neurons that are tuned for increasingly off-vertical orientations. When a central stimulus is embedded in a surrounding stimulus, neurons that would typically fire for a central stimulus with an orientation similar to the surround are suppressed while neurons tuned to substantially different orientations are left unaffected. As a result, when a vertical stimulus is embedded in a surround that is tilted slightly leftward, say, neurons that tend to prefer slightly right-tilting stimuli in the center are suppressed while other neurons that respond to central stimuli that are tilted in the other direction from the surround are unaffected, left to murmur at their normal level. The hush of the "slightly left-tilted"-coding neurons together with the continuing murmurs of the neurons that code for "slightly right-tilted" orientations mean that the whole population of neurons that are responding to the central stimulus behaves approximately as it would to an isolated stimulus that is tilted slightly rightward. This population-level response aligns well with perception, since one in fact perceives the central vertical stimulus as tilted slightly away from its surround. It is reasonable to conclude that these modulations of neural activity give rise to distinct types of perceptual representations of the central contours: $\mathbf{c_{0^\circ}}$, say, in the context of representing a parallel surround, and $\mathbf{c_{2^\circ}}$ in the context of representing a 20° left-tilted surround.

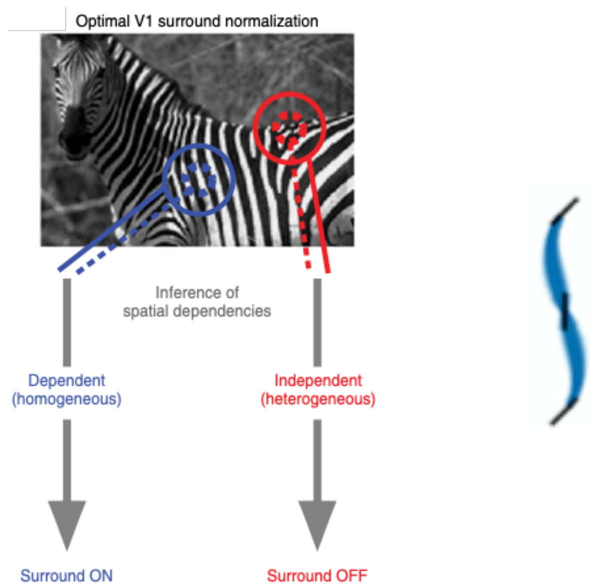
The strength of tilt effects depends significantly on the strength of the perceptual evidence that the central line or grating and the surrounding pattern belong to a common contour or pattern. Tilt effects become substantially weaker when the central stimulus is perceptually segregated from the surrounding pattern, for example when the central contours appear to be at different depths, to have different levels of contrast, or to be too misaligned (Qiu et al., 2013). The degree to which tilt representations suppress each other has also been shown to increase with the likelihood that the represented contours would be part of the same object in a typical scene (Schwartz et al., 2009, Coen-Cagli et al., 2012). Further, a number of models of contour integration predict the tilt effect, suggesting that the effect is intimately related to the principles governing how representations of contour segments combine into a representation of a cohesive contour (Kapadia et al., 2000, Schwartz et al., 2005, Keemink et al., 2018, Linsley et al., 2020).

4.2.2 Compositional Analysis

The sort of analysis I will propose is that, abstracting away from represented positions, a representation of the form $\ulcorner \mathbf{c_{0^\circ}} \circ \mathbf{c_{-20^\circ}} \urcorner$, representing a vertical segment connected to a segment that is tilted 20° to the left, generally will have a weaker binding strength than representations of the form $\ulcorner \mathbf{c_{2^\circ}} \circ \mathbf{c_{-20^\circ}} \urcorner$, representing a 2° right-tilted segment connected to a 20° left-tilted surround. So, given that one is representing the surround as tilted 20° to the left, there is a structural preference to represent the center as tilted slightly to the right. In order to counteract this structural preference and represent the central contours as vertical, the physical contours would in fact have to be tilted slightly toward the surround.

For a general account, I draw on Schwartz et al. (2009)'s model of tilt effects (see also Schwartz et al., 2007, Coen-Cagli et al., 2012, 2015). According to this model, the representations of the different parts of a common contour or pattern must be *normalized*, so that there is as little statistical coordination between the representations as possible. Neighboring segments along a contour tend to be highly correlated in their orientations. If one segment is vertical, the next connecting segment is likely to be close to vertical and unlikely to be horizontal. While some degree of one segment's orientation will be predictable from the neighboring segment's orientation, some degree will not be. Representations of the segments are "normalized" with respect to each other to the extent that the representation of the one is conditionally independent of the representation of the other, given that they both represent connected parts of a common contour. What this means is that the representations carry as little redundant information about each other as possible, given that they are representing the same contour. Whereas many models of perception entail that the representation of a given item must be normalized with respect to the representation of other items, a crucial component of Schwartz et al.'s model is the assumption that representations should only be normalized with respect to representations of the *same structure*. So, the constituents of a complex representation should be normalized with respect to each other but not with respect to other unrelated representations (Figure 4a). As a structural constraint, a set of representations is more able to combine the more the constituents are normalized with respect to each other—that is, the more they represent non-redundant, distinctive features of their respective targets (see also Barlow, 2001).

Under conditions of uncertainty, the normalization constraint will usually bias the visual system toward exaggerating the contrastive features of an object. Suppose the visual system treats two sensory cues as arising from different parts of the same object, which is estimated to be largely uniform. Suppose these cues indicate a small but non-accidental difference in orientation between two parts of the object—a difference that cannot be attributed merely to noise. The normalization constraint gives greater weight to combining representations according to which these parts really differ than it does to combining representations according to which the parts are redundant. Hence the "repulsion" in the perceived orientations of the center and surround when these are represented



(a) *(b)*

Figure 4: (a) Representations of parts of the same contour or pattern should be normalized with respect to each other but not with respect to representations of independent objects (from Coen-Cagli et al., 2015, p. 1651). (b) Representing the central bar as connecting the flankers by the smoothest curve (indicated here in blue) requires representing the central bar as tilted slightly away from the flankers (from Keemink and van Rossum, 2016, p. 168).

as parts of the same object and when they give rise to tilt cues that differ slightly but significantly.[12]

In fact, a number of models suggest that a bias against combining constituents that represent orientations that are either too similar or too dissimilar is derivable from something like the good continuation constraint described in the previous section (Kapadia et al., 2000, Schwartz et al., 2005, Keemink and van Rossum, 2016). The alignment between contours is a function of both the position and orientation of those contours (Field et al., 1993). Except when two contour segments form a straight line, the smoothest continuation from one to the next will typically require that they have different orientations, depending on their respective positions. While one segment is unlikely to differ radically in orientation from the last, they are unlikely to be identical either. In the specific situations that give rise to the repulsive tilt effect, "the smoothest continuations of the surround elements tilts the percept away from the surround and, in special cases, attracts it" (Keemink and van Rossum, 2016, p. 171; Figure 4b). The specific bias against combinations of

[12]A similar explanation can be given for why there is a slight attractive effect when the target and surround have very different orientations. In this case, the normalization requirement privileges the representation of non-accidentally *similar* orientations in a stimulus for which heterogeneity is the norm (see Schwartz et al., 2009).

similar representations may therefore be part and parcel of a constraint on representing smoothly aligned contours.

4.2.3 Representational Function of the Tilt Effect

It is natural to treat tilt effects as biasing one toward inaccurate estimates of the orientations of things, over-estimating differences at the cost of accuracy. It is true that tilt effects sacrifice perfect accuracy on the rare occasions where it might be possible. But we perceive under conditions of uncertainty. One way to reduce uncertainty is to rely on the relationships between items in the world. A self-standing contour segment is a rarefied thing. Contour segments are almost always local parts of some larger, smoothly connected structure. So even if there is evidence that a particular segment is vertical, say, evidence that the segment connects to surrounding segments may call for representing the central segment in a way that makes most sense of that segment's position in the overall structure being represented. In this case, that means representing what is distinctive about the item in a way that best integrates into a representation of a smoothly continuing contour. A system that does not exhibit tilt effects, because it does not take into account surrounding information and instead treats all contours as independently "shrink-wrapped," to borrow a phrase form Wijntjes and Rosenholtz (2018), would likely be less accurate in representing whole contours and the distinctive features of their parts (see, for example, Linsley et al., 2020). Tilt effects exemplify a sensitivity to and capitalization on the structure of items in the world in order to compose representations that are reliably, if only approximately, accurate.

Moreover, the normalization constraint demands compositional structures in which each constituent representation of a part is as insulated as possible from the potential error of other constituents. A consequence of normalization is that the representation that is formed for one part of an item carries more content that is distinctive to that part while implying less about the features of other parts. This means that the inaccuracy or unreliability of one constituent need not imply the inaccuracy or unreliability of the others. When these representations combine into a whole, the possibilities for error are thereby regimented (Attneave, 1954, Barlow, 2001). The complex representation that these constituents form has the greatest chance of being accurate despite variability in the accuracy of its parts. Tilt effects arise from biases for combinations of representations that make content about each local part of an object reliably accurate and independently extractable within a reliably accurate representation of the whole object.

4.3 Reference Frame Effects

So far I have argued that context can influence perception by regulating either the structural relationships between representations or through selecting the types of representations that enter into such relationships. In each of these cases, a compositional analysis requires giving up the principle that perceptual representations can combine more or less without substantive restriction. I will motivate a different type of analysis of the effect that a *frame of reference* has on the perception of an item's orientation. Figure 3 offers one illustration. As Attneave (1968) pointed out, an equilateral triangle might be seen as "pointing" in any of three directions, corresponding to its three corners, but it is never seen as pointing in more than one direction at a time. Context heavily influences which of these orientations is attributed to the triangle. The very same triangle in Figure 3 is seen as pointed to the right when embedded in a horizontal rectangle, while it is seen as pointed to the upper left when embedded in an obliquely oriented rectangle.

Palmer (1980) demonstrated that, other things equal, if a figure is ambiguous with respect to its orientation, one tends to see it as having the orientation that is closest to being either parallel or perpendicular to a whole configuration of which the figure is perceived as a part. In kind with the previous analysis of tilt effects, one could propose that reference frame effects merely consist in the structural selection of different representations of the triangle's orientation. Suppose that the triangle admits of a set of potential representations, $\mathtt{t}_{\theta}$, $\mathtt{t}_{\theta'}$, and $\mathtt{t}_{\theta''}$, each of which attributes a different absolute orientation. On this account, if the rectangle is represented as oriented a certain way, then the representation of the rectangle is most able to combine with whichever triangle representation assigns an orientation that is closest to parallel or perpendicular with the rectangle's represented orientation. Against this hypothesis, a significant body of literature suggests that one tends to represent the triangles in Figure 3 as having the same *relative orientation* in each case: as *oriented parallel to* ___. I will suggest that this representation does not determine a complete accuracy condition. Much like the adjectival modifier "former," the representation does not itself characterize an item in a way that can be evaluated as accurate or inaccurate. Instead, the representation of relative orientation constitutively functions to combine with a representation of an object that serves as a reference frame, so as to compose a more complex representation of the triangle as oriented parallel to that particular object. It is this more complex representation that accurately or inaccurately represents the orientation of the triangle relative to the orientation of the rectangular reference frame.

4.3.1 Evidence for the Representation of Relative Orientation

The claim that we perceptually represent things as having orientations relative to some object or objects that serve as an extrinsic frame of reference—for example, that one perceptually represents the triangle as oriented parallel to the surrounding rectangle—has substantial theoretical precedent, playing an especially central role in theories of shape representation, object recognition, and motion perception (Marr and Nishihara, 1978, Palmer, 1989, Xu et al., 2017, Lauffs et al., 2019). One source of support for the claim comes from patients with "object-based" visual neglect. Visual neglect can be a symptom of cortical lesions due to stroke and is normally characterized by an inability to attend to certain areas in the scene. For example, many neglect patients have difficulty attending to the left of the visual field with respect to either their line of sight, the orientation of their head, or the orientation of their torso. Intriguingly, many studies have reported cases of *object-based neglect*. For example, Tipper and Behrmann (1996) showed patients a horizontal "dumbbell" figure consisting of two textured circles connected by a straight line and asked them to detect whether one of the circles was a predefined target. Most subjects had difficulty identifying the target when it was on the left. In the crucial condition, after appearing on screen, the dumbbell rotates 180° in full view. If one perceives the position of a part of the dumbbell with respect to the whole object, then whichever circle one perceives to be on the right of the dumbbell prior to rotation should continue to be perceived as on the right of the dumbbell after rotation, even though it is now located on the left side of the screen. Tipper and Behrmann found that after rotation, some patients consistently were able to detect the right part of the dumbbell, even though it was now located on the left side of the screen. However, they continued to have difficulty detecting the left part of the dumbbell, though its final position was on the right side of the screen. This suggests that the patients were neglecting parts in the left half *of the object* and not just in the left half of the visual field. In order to explain this, it seems that we must posit that the same circle is represented the same way before and after orientation, as *left of y*, although the orientation of the object, y, changes through the rotation.

While much of the literature on object-based neglect focuses on the representation of positions of items within a reference frame, similar points can be made for the perceived orientation of items within a reference frame (see Robertson, 2004). Take, for example, McCloskey and his colleagues' analyses of the difficulties that neurotypical subjects, children, and patients with developmental deficits have in discriminating objects from their left-right reflected "mirror-images" (McCloskey et al., 2006, Gregory and McCloskey, 2010, Gregory et al., 2011). Whether two stimuli are treated as mirror images depends on what one represents as the left and right parts of those stimuli, and hence how one perceives the stimuli as oriented. McCloskey and his colleagues find that we do not just mistake stimuli that are reflected about their own axis, as with "" and " in which the bump is reflected around the stem; we also mistake stimuli that are reflected about an external axis, such as "" and ," in which the whole figure is left-right reflected with respect to the orientation of the page. According to their "coordinate-system orientation representation" hypothesis, "an object's orientation relative to an extrinsic reference frame is represented by specifying the relationship between the axes of the object-centered frame and those of the extrinsic frame" (Gregory and McCloskey, 2010, p. 124).[13]

Visual search also appears to be sensitive to an item's orientation relative to an external reference frame. Treisman and Gormican (1988) demonstrated that an oblique line is especially salient in a field of vertical lines. An oblique line "pops out" from the vertical lines, phenomenologically

[13] In fact, these authors argue that the variety of mirror-image confusions requires representations of frame-relative orientation themselves to be complex: "object-orientation representations are compositional—that is, the representations are composed of multiple components, each of which represents a different aspect of an object's orientation," including constituents that respectively specify the polarity, direction, and magnitude of the item's orientation relative to its extrinsic frame of reference (Gregory and McCloskey, 2010, p. 127). Neander (2017, Ch. 1) discusses these results, but with a somewhat different purpose.

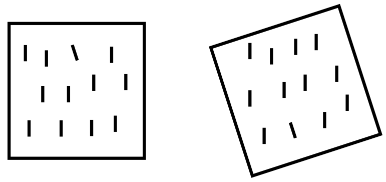


Figure 5: An oblique line "pops out" from vertical distractors in an upright frame but not in a frame that is tilted parallel to the line.

speaking, and one's ability to detect the oblique line does not diminish substantially with greater numbers of vertical distractors. Interestingly, this pop-out effect is not symmetrical: a vertical target does not pop out in a field of oblique distractors. However, that vertical target will pop out if it is presented in a tilted frame so that it is oblique *with respect to the frame* and the distractors are vertical *with respect to the frame* (Figure 5; May and Zhaoping, 2009, Marendaz, 1998).

4.3.2 Compositional Analysis

An item can be represented as having a given relative orientation in a number of different contexts. The triangles in Figure 3 are each represented as parallel to their reference frame, though they are represented as having different reference frames. What remains constant across these two contexts is the representation of the triangle as oriented parallel to an item y, where y is not specified by the representation of the triangle's relative orientation. The difference in the perceived orientations of the triangles is a result of combining the representation of the triangle's relative orientation with the representations of a specific object that serves as the reference item, y, with respect to which the relative orientation of the triangle determines an absolute orientation in the visual field.

The representation of relative orientation is in a sense an "incomplete" or "unsaturated" representation. I will assume that the representation of an item's relative orientation cannot occur except in combination with another representation that specifies its reference frame. The same representation of the triangle as oriented parallel to its reference frame can occur in combination with different representations of reference frames; but the former cannot occur in the absence of any separate representation of an extrinsic reference frame's orientation, shape, or other features.

When the representation of relative orientation is combined with a representation of another item, the result is a complex representation in which the value of y in the content of the former representation is given by (bound to) the object represented by the latter.

Let $\alpha_{\theta\rightarrow}$ be a representation of an item's orientation θ as a function of some unspecified reference frame y. We can give the conditions under which this representation of the triangle is *satisfied by*, or perfectly accurate *relative to*, a reference frame y:

> If $\alpha_{\theta\rightarrow}$ is a representation of x, then it is perfectly accurate relative to y iff x is oriented at an angle θ° to the orientation of y.

This representation cannot be evaluated as accurate or inaccurate full stop, since the reference item y is not actually given. Rather the representation functions to be combined with another representation and its content functions to be integrated with the content of that other representation.

Let $\ulcorner \alpha_{\theta\rightarrow} \odot \beta_{\theta'} \urcorner$ denote the structural relation between a representation $\alpha_{\theta\rightarrow}$ of an item's relative orientation and a representation $\beta_{\theta'}$ of another item that serves as its reference frame. Idealizing, and abstracting away other details such as represented position, motion, and so on, the structural constraint on this mode of combination has something like the following form:

Framing: (6) (a) $\alpha_{\theta\rightarrow}$ can only occur in combination with another representation, $\ulcorner(\alpha_{\theta\rightarrow} \odot \beta_{\theta'})\urcorner$.

(b) The binding strength of $\ulcorner(\alpha_{\theta\rightarrow} \odot \beta_{\theta'})\urcorner$ is proportional to how close θ is to a cardinal orientation ($0^\circ, 90^\circ, 180^\circ$ and 270°).

The principle of semantic composition for $\ulcorner(\alpha_{\theta\rightarrow} \odot \beta_{\theta'})\urcorner$ sets the reference orientation for $\alpha_{\theta\rightarrow}$ to the orientation of the object represented by $\beta_{\theta'}$:

Framing: (7) $\ulcorner(\alpha_{\theta\rightarrow} \odot \beta_{\theta'})\urcorner$ is perfectly accurate iff

(a) if $\beta_{\theta'}$ is a representation of y, then $\alpha_{\theta\rightarrow}$ is perfectly accurate relative to y (that is, $\alpha_{\theta\rightarrow}$ is perfectly accurate if the free variable in its accuracy condition is bound to the value of y), and

(b) $\beta_{\theta'}$ is perfectly accurate.[14]

[14] These clauses can be revised to accommodate the case in which multiple representations combine under a common reference frame: $\ulcorner(\{\alpha_{\mathbf{i}\theta\rightarrow}, \ldots, \alpha_{\mathbf{n}\theta'\rightarrow}\} \odot \beta_{\omega})\urcorner$.

Suppose $\mathbf{t_{0^\circ\rightarrow}}$ is the representation of the triangle as oriented 0° from the orientation of an unspecified item y and $\mathbf{r_{90^\circ}}$ is a representation of a rectangle as oriented 90°. Then $\mathbf{t_{0^\circ\rightarrow}} \odot \mathbf{r_{90^\circ}}$ is perfectly accurate just in case the triangle is oriented 0° from the rectangle and the rectangle is oriented 90°. The accuracy of this representation therefore depends on two things: the triangle's orientation relative to the rectangle and the orientation of the rectangle. The whole representation might be accurate with respect to the one but inaccurate with respect to the other.

The theory sketched in Section 2 characterizes the constituents of perceptual representations as semantically uniform in the sense that each has the same type of semantic role: each determines a complete accuracy condition. By contrast, while the representation of the triangle as oriented parallel to its extrinsic reference frame does carry content about the triangle, it does not determine a complete accuracy condition; its content functions to integrate with the content of another representation that specifies the reference frame. Perceptual representations of various aspects of a scene may therefore have different types of semantic values (cf. Fodor, 2008, Shea et al., 2017). In fact, this analysis entails that constituents of arbitrary complexity can be semantically incomplete. One can get multiple embeddings of reference frames. For example, if $\mathbf{a_{0^\circ\rightarrow}}$, $\mathbf{b_{90^\circ\rightarrow}}$, and $\mathbf{c_{30^\circ}}$ represent a, b, and c respectively, then the combination $(\mathbf{a_{0^\circ\rightarrow}} \odot \mathbf{b_{90^\circ\rightarrow}}) \odot \mathbf{c_{30^\circ}}$ is perfectly accurate just in case a is oriented in the same direction as b, b is oriented 90° from c, and c is oriented 30° (see Watt, 1990, Baylis and Driver, 1993). The complex constituent, $(\mathbf{a_{0^\circ\rightarrow}} \odot \mathbf{b_{90^\circ\rightarrow}})$, does not determine a complete accuracy condition. While it may accurately represent a's orientation relative to b, it functions to combine with a representation of an item relative to which b's orientation is represented.

4.3.3 Representational Function of Reference Frame Effects

As Marr and Nishihara (1978) noted, there is an ecological rationale for representing position, orientation, and other spatial features in a compositional, frame-relative fashion. The spatial relationships between parts of an object offer significant cues to the identity and category of the object, not to mention its functional significance. If one needs to tell a wolf from a fox, the spatial relationships between the head, torso, limbs, and tail are a dead giveaway, but whether the creature's tail is vertical or horizontal relative to one's visual field is incidental. If one needs to tell an aggressive wolf from an ambivalent one, what matters is the orientation of the tail relative to the wolf's torso, not its orientation relative to oneself. By representing the positions and orientations of items relative to the larger objects or configurations of which those items are parts, the visual system makes these critical relationships explicit.

Representations of frame-relative position and orientation are also more likely to be reliably accurate than representations that merely attribute absolute orientations to all items. Normally, the different features in a scene are not positioned entirely independently of each other. The position, orientation, and motion of an object constrains the positions, orientations, and motions of its parts. The wolf's tail is unlikely to be very far separated from the wolf's head, no matter where the wolf is in the visual field. As with normalization, representing the positions and orientations of items relative to the objects or configurations of which they are part regiments the risk of error. Because the range of different positions and orientations that the wolf's legs can have relative to its body is far smaller than the range of different positions and orientations that they can have relative to the viewer or the entire scene, there is far less risk of inaccuracy in representing the legs relative to the body than there is in representing their positions and orientations relative to oneself or the whole scene. Even if one is substantially wrong about where the whole is, one can still be substantially right about its parts and their relations.[15]

5 Context and Objectivity

Each of the analyses that I advocated in the previous section locates the source of contextual influence in a different aspect of the compositional edifice: in the structural relationships between representations, in the selection of constituents that will best fit into such relationships, or in semantic operations over incomplete contents. Even though different context effects require different analyses, they have a common representational function. I argued in each case that the context effects functioned to facilitate the composition of reliably accurate representations, given an uncertain but

[15] None of the foregoing is meant to suggest that there is no perceptual representation of the triangle *per se* as having an absolute orientation, only that the triangle is *at least* represented as having a frame-relative orientation.

structured environment. I suggest that this is a general function of perceptual context effects and characteristic of the compositional grammar of perception.

To briefly review: (1) Representations of distant, substantially misaligned contour segments cannot combine into a representation of a whole contour, unless they are also combined with representations of connecting segments. This structural principle reflects the environmental regularity that two distant, disconnected, substantially misaligned edges are extraordinarily unlikely to be part of the same contour. (2) Representations of contour segments selectively combine with representations of contour segments with dissimilar, but aligned orientations. This principle capitalizes on the regularities that hold among the different parts of a contour in order to reliably represent the whole contour, as well as the genuinely distinctive features of the different parts of that contour. (3) By representing features of an item relative to an object or configuration that serves as an extrinsic reference frame, the representation of the whole and its parts can be insulated somewhat from uncertainty about the rest of the scene. This is because the internal structure of an object or configuration tends to be more stable than the relations between disparate objects or configurations. The constituent representation of an item's relative orientation is therefore able to contribute content about a stable feature of the item to the content of the whole percept.

It is widely recognized that perception capitalizes on systematic patterns in the world in order to accurately represent items across a variety of settings in which they can appear. The claim here is that substantive environmental patterns are recapitulated in the compositional grammar of perception, namely the ways perceptual representations can and cannot combine and the semantic import of those combinations. The combinability of representations in different contexts of embedding reflects the varying relationships that the represented features have across different settings. The semantic contributions of perceptual representations encodes the stable role that the represented features play in different settings. Insofar as a perceptual state is structurally possible, it is unlikely to represent something that is highly abnormal in the perceiver's environment—a contour the parts of which are randomly scattered across the scene, for example. The structural possibility of a perceptual state is a guide to its objective plausibility.

The combinatorial and semantic principles in perception capitalize on objective patterns in the world. But the world cannot just imprint these principles on the mind through sensory signals. The retinal image underdetermines which edges belong to the same object, for example. Regularities about how object contours hang together are not inherent in the surface patterns in the sensory stimulus. Instead, the mind has to develop, through evolution and learning, its own internal constraints on how representations can go together so as to best reflect how aspects of the world go together (Geisler and Diehl, 2002, Kemp and Tenenbaum, 2008). These internal constraints must be enforced in the constructive parsing of the unstructured image. If the compositional principles in perception are biased against the composition of wildly inaccurate representations, then the free combinability and semantic uniformity of representations should be the exception rather than the rule in perception. In a structured but uncertain world, reliable accuracy requires that representations not be structurally inert islands. The representation of one feature should not be "shrink-wrapped" from the representation of other relevant features. Instead, there should be substantive combinatorial constraints and differentiation in semantic roles that reflect the patterned relationships between the represented features in the world.

I propose, then, that the context effects discussed here, and perceptual context effects in general, are the product of what Kemp and Tenenbaum (2008) call "domain-specific constraints on the structure of mental representations." The constraints mentioned above function to ensure that the visual system cannot compose representations of situations that are radically improbable in the perceiver's normal environment. By contrast, many other familiar representational schemes do not have this character. The well-formedness of a logical formula, not to mention a sentence of English, does not typically provide any guarantee to its truth or plausibility. Even contradictory statements can be well-formed! These schemes give wide syntactic permission to represent the ridiculous and even impossible. One explanation for the central role of context effects in perception is that the compositional principles of perception—the very *forms* of perceptual representations—aim at the composition of reliably accurate representations of the perceiver's normal environment .

6 Conclusion

A compositional semantics of perception explains how the accuracy conditions of perceptual representations derive from the contents of their constituent parts and the way these parts are structurally related. Such a theory is of broad value for understanding the nature, grounds, and epistemic role of perceptual content. However, the explanatory power and generality of many existing accounts of the semantics of perception are unclear. As an illustration, I outlined a basic semantics for a rarefied set of percepts. This account had two notable features: it characterized perceptual representations of the different aspects of a scene as semantically uniform, each determining a complete accuracy condition, and as combinable without substantive restriction. I pointed to the limitations of this account, and others like it, when it comes to the sort of "holistic" perceptual phenomena that has been at the center of research on perceptual organization for over a century. In the first place, one can represent some sets of items, but not others, as having certain higher-order features. Associated with the perception of higher-order, emergent features are context effects, in which the way one aspect of a scene is perceived depends on the way other aspects of the scene are perceived. Perceptual context effects are ubiquitous and central to perception. An adequate semantics of perception must accommodate these phenomena. However, formulating such a semantics is a non-trivial task and some have doubted that it is a feasible one.

I argued that characteristic examples of context-sensitivity in perception are driven by an underlying compositional grammar. But this grammar is not one in which perceptual representations are semantically uniform and freely combinable. Why should perception lack the semantic uniformity and free combinability implicit in the basic semantics with which we began? I suggested that context effects are closely tied to the objectivity of perception. The grammar of perception functions to ensure the composition of reliably accurate percepts in an uncertain but structured world. The combinability of representations reflects the normal combinability of the represented features in the world and the semantic value of a representation reflects the stable role of the represented feature in the different surroundings in which it occurs. Context effects arise when order is brought to one's internal perceptual representations, an order which capitalizes on the natural order in the features one perceptually represents.

Compositional semantics offers a framework for understanding the complex ways that mental states constitutively function to fit together to represent objective features of the world. An adequate understanding of the nature of a given perceptual representation requires an understanding of its systematic, nuanced relationships to its constituents and to representations of which it is a constituent. It remains to be seen how far this semantic program can be pushed for perception—not just for representations in vision, but also for those in and across other modalities. One aim of this discussion has been to show that the issues involved in working out an adequate semantics of perception are far richer and more varied than has generally been appreciated. Even setting foundational questions to the side, the core explanatory challenge of accounting for the diverse phenomena of perception is daunting. These phenomena regularly resist straightforward semantic analysis. Still, I think a constructive optimism toward a compositional semantics of perception is warranted. Despite their different analyses, the phenomena discussed here share a representational function that is plausibly common throughout perception. The compositional grammar of perception, context effects and all, constitutes a framework for supporting the objectivity of perception.

References

Ankrum, Corlene and Palmer, John. (1991). Memory for objects and parts. *Perception & Psychophysics 50*(2), 141–156.

Attneave, Fred. (1954). Some informational aspects of visual perception. *Psychological Review 61*(3), 183–193.

Attneave, Fred. (1968). Triangles as ambiguous figures. *The American Journal of Psychology 81*(3), 447–453.

Balog, Katalin. (2009). Jerry Fodor on non-conceptual content. *Synthese 170*(2), 311–320.

Barenholtz, Elan and Feldman, Jacob. (2003). Visual comparisons within and between object parts: evidence for a single-part superiority effect. *Vision Research 43*(15), 1655–1666.

Barlow, Horace. (2001). Redundancy reduction revisited. *Network: Computaiton in Neural Systems 12*(3), 241–253.

Baylis, Gordon C. and Driver, Jon. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance 19*(3), 451–470.

Beck, Jacob. (2019). Perception is analog: The argument from Weber's law. *The Journal of Philosophy 116*(6), 319–249.

Brady, Timothy F. and Alvarez, George A.. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision 15*(6), 1–24.

Brady, Timothy F., Konkle, Talia, and Alvarez, George A.. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision 11*(5), 1–34.

Brunswik, Egon and Kamiya, Joe. (1953). Ecological cue-validity of 'proximity' and of other gestalt factors. *The American Journal of Psychology 66*(1), 20–32.

Burge, Tyler. (2005). Frege on truth. In *Truth, Thought, Reason: Essays on Frege*. Oxford University Press.

Burge, Tyler. (2010). *Origins of Objectivity*. New York: Oxford University Press.

Camp, Elisabeth. (2007). Thinking with maps. *Philosophical Perspectives 21*(1), 145–182.

Camp, Elisabeth. (2018). Why maps are not propositional. In Alex Grzankowski and Michelle Montague (Eds.), *Non-Propositional Intentionality*, pp. 19–45. Oxford: Oxford University Press.

Clark, Austen. (2004). Feature-placing and proto-objects. *Philosophical Psychology 17*(4), 443–469.

Clifford, Colin W.G.. (2014). The tilt illusion: Phenomenology and functional implications. *Vision Research 104*, 3–11.

Coen-Cagli, Ruben, Dayan, Peter, and Schwartz, Odelia. (2012). Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS Computational Biology 8*(3), e1002405.

Coen-Cagli, Ruben, Kohn, Adam, and Schwartz, Odelia. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience 18*(11), 1648–1655.

Cummins, Robert. (1996). *Representations, Targets, and Attitudes*. MIT Press.

Devitt, Michael. (2006). *Ignorance of Language*. New York: Oxford University Press.

Elder, James H. and Goldberg, Richard M.. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision 2*(4), 324–353.

Elder, James H., Oleskiw, Timothy D., and Fruend, Ingo. (2018). The role of global cues in the perceptual grouping of natural shapes. *Journal of Vision 18*(12), 1–21.

Feldman, Jacob, Singh, Manish, and Froyen, Vicky. (2014). Perceptual grouping as Bayesian mixture estimation. In Sergei Gepshtein, Larry Maloney, and Manish Singh (Eds.), *The Oxford Handbook of Computational Perceptual Organization*. Oxford University Press.

Field, David J., Hayes, Anthony, and Hess, Robert F.. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research 33*(2), 173–193.

Fodor, Jerry. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, Jerry. (2008). *LOT 2: The Language of Thought Revisited*. New York: Oxford University Press.

Fodor, Jerry and Pylyshyn, Zenon. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition 28*(1–2), 3–71.

Gauker, Christopher. (2012). Perception without propositions. *Philosophical Perspectives 26*(1), 19–50.

Geisler, Wilson S. and Diehl, Randy L.. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357*(1420), 419–448.

Geisler, Wilson S., Perry, J.S., Super, B.J., and Gallogly, D.P.. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research 41*(6), 711–724.

Geisler, Wilson S. and Super, Boaz J.. (2000). Perceptual organization of two-dimensional patterns. *Psychological Review 107*(4), 677–708.

Geman, Stuart, Potter, Daniel F., and Chi, Zhiyi. (2002). Composition systems. *Quarterly of Applied Mathematics 60*(4), 707–736.

Gilbert, Charles D. and Wiesel, Torsten N.. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research 30*(11), 1689–1701.

Golomb, Julie D., Kupitz, Colin N., and Thiemann, Carina T.. (2014). The influence of object location on identity: A "spatial congruency bias". *Journal of Experimental Psychology: General 143*(6), 2262–2278.

Gregory, Emma, Landau, Barbara, and McCloskey, Michael. (2011). Representation of object orientation in children: Evidence from mirror-image confusions. *Visual Cognition 19*(8), 1035–1062.

Gregory, Emma and McCloskey, Michael. (2010). Mirror-image confusions: Implications for representation and processing of object orientation. *Cognition 116*(1), 110–129.

Hess, Robert F., May, Keith A., and Dumoulin, Serge O.. (2015). Contour integration: Psychophysical, neurophysiological, and computational perspectives. In Johan Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization*, Chapter 10, pp. 189–206. Oxford University Press.

Higginbotham, James. (1986). Linguistic theory and Davidson's program in semantics. In Ernest LePore (Ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, pp. 29–48. Basil Blackwell.

Higginbotham, James. (2003). Conditionals and compositionality. *Philosophical Perspectives 17*(1), 181–194.

Hintikka, Jaakko. (1983). Theories of truth and learnable languages. In *The Game of Language: Studies in Game-Theoretical Semantics and Its Applications*, Chapter 10, pp. 259–292. Boston: D. Reidel Publishing Company.

Janssen, Theo M.V.. (2011). Compositionality. In Johan van Benthem and Alice ter Meulen (Eds.), *Handbook of Logic and Language* (2nd ed.)., Chapter 10, pp. 495–553. Elsevier.

Kapadia, Mitesh K., Ito, Minami, Gilbert, Charles D., and Westheimer, Gerald. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert Monkeys. *Neuron 15*, 843–856.

Kapadia, Mitesh K., Westheimer, Gerald, and Gilbert, Charles D.. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology 84*(4), 2048–2062.

Keane, Brian P.. (2018). Contour interpolation: A case study in modularity of mind. *Cognition 174*, 1–18.

Keemink, Sander W., Boucsein, Clemens, and van Rossum, Mark C. W.. (2018). Effects of v1 surround modulation tuning on visual saliency and the tilt illusion. *Journal of Neurophysiology 120*(3), 942–952.

Keemink, Sander W. and van Rossum, Mark C.W.. (2016). A unified account of tilt illusions, association fields, and contour detection based on elastica. *Vision Research 126*, 164–173.

Kellman, Philip J and Shipley, Thomas F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology 23*(2), 141–221.

Kemp, C. and Tenenbaum, J. B.. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences 105*(31), 10687–10692.

Kempgens, Christian, Loffler, Gunter, and Orbach, Harry S.. (2013). Set-size effects for sampled shapes: Experiments and model. *Frontiers in Computational Neuroscience 7*(67), 1–18.

Kimchi, Ruth, Yeshurun, Yaffa, Spehar, Branka, and Pirkner, Yossef. (2015). Perceptual organization, visual attention, and objecthood. *Vision Research 126*, 34–51.

Kulvicki, John V.. (2015). Analog representation and the parts principle. *Review of Philosophy and Psychology 6*(1), 165–180.

Lande, Kevin J.. (2021). Mental structures. *Noûs 55*(3), 649–677.

Larson, Richard and Segal, Gabriel. (1995). *Knowledge of Meaning: An Introduction to Semantic Theory*. Cambridge, MA: MIT Press.

Lauffs, Marc M., Choung, Oh-Hyeon, Ögmen, Haluk, Herzog, Michael H., and Kerzel, Dirk. (2019). Reference-frames in vision: Contributions of attentional tracking to nonretinotopic perception in the ternus-pikler display. *Journal of Vision 19*(12), 7.

Li, Zhaoping. (1999). Contextual influences in v1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences 96*(18), 10530–10535.

Linsley, Drew, Kim, Junkyung, Ashok, Alekh, and Serre, Thomas. (2020). Recurrent neural circuits for contour detection. In *International Conference on Learning Representations*.

Marendaz, Christian. (1998). Nature and dynamics of reference frames in visual search for orientation: Implications for early visual processing. *Psychological Science 9*(1), 27–32.

Marr, David and Nishihara, Herbert Keith. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences 200*(1140), 269–294.

Matthen, Mohan. (2004). Features, places, and things: Reflections on austen clark's theory of sentience. *Philosophical Psychology 17*(4), 497–518.

Matthen, Mohan. (2005). *Seeing, Doing, & Knowing: A Philosophical Theory of Sense Perception*. Oxford: Oxford University Press.

Matthen, Mohan. (2014). Image content. In *Does Perception Have Content?* Oxford University Press.

May, Keith A. and Zhaoping, Li. (2009). Effects of surrounding frame on visual search for vertical or tilted bars. *Journal of Vision 9*(13), 1–19.

McCloskey, Michael, Valtonen, Jussi, and Cohen Sherman, Janet. (2006). Representing orientation: A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology 23*(5), 680–713.

Nanay, Bence. (2020). Perceiving indeterminately. *Thought: A Journal of Philosophy 9*(3), 160–166.

Neander, Karen. (2017). *A Mark of the Mental: In Defense of Informational Teleosemantics*. Cambridge, MA: The MIT Press.

Palmer, Stephen E.. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology 9*(4), 441–474.

Palmer, Stephen E.. (1980). What makes triangles point: Local and global effects in configurations of ambiguous triangles. *Cognitive Psychology 12*(3), 285 – 305.

Palmer, Stephen E.. (1989). Reference frames in the perception of shape and orientation. In Bryan E. Shepp (Ed.), *Object Perception: Structure & Process*, Volume Ballesteros, Soledad, Chapter 5, pp. 121–163. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Palmer, Stephen E. and Beck, Diane M.. (2007). The repetition discrimination task: An objective method for studying perceptual grouping. *Perception & Psychophysics 69*(1), 68–78.

Partee, Barbara. (2004). Compositionality. In *Compositionality in Formal Semantics*, pp. 153–181. Blackwell.

Partee, Barbara H.. (2007). Compositionality and coercion in semantics: The dynamics of adjective meaning. In Gerlof Bouma, Irene Krämer, and Joost Zwarts (Eds.), *Cognitive Foundations of Interpretation*, pp. 145–161. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

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

Pomerantz, James R. and Cragin, Anna I.. (2015). Emergent features and feature combination. In Johan Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization*, pp. 88–107. Oxford: Oxford University Press.

Pomerantz, James R. and Pristach, Edward A.. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance 15*(4), 635–649.

Qiu, Cheng, Kersten, Daniel, and Olman, Cheryl A.. (2013). Segmentation decreases the magnitude of the tilt illusion. *Journal of Vision 13*(13), 19–19.

Quilty-Dunn, Jake. (2019). Perceptual pluralism. *Nous Advanced online publication*.

Quine, Willard Van Orman. (1986). *Philosophy of Logic* (2nd ed.). Harvard University Press.

Rescorla, Michael. (2009). Cognitive maps and the language of thought. *British Journal for the Philosophy of Science 60*(2), 377–407.

Rescorla, Michael. (2019). The language of thought hypothesis. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. URL = https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=language-thought.

Robertson, Lynn C.. (2004). *Space, Objects, Minds, and Brains*. Essays in Cognitive Psychology. Psychology Press.

Schellenberg, Susanna. (2018). *The Unity of Perception: Content, Consciousness, Evidence*. New York: Oxford University Press.

Schwartz, Odelia, Hsu, Anne, and Dayan, Peter. (2007). Space and time in visual context. *Nature Reviews Neuroscience 8*(7), 522–535.

Schwartz, Odelia, Sejnowski, Terrence J., and Dayan, Peter. (2005). A Bayesian framework for tilt perception and confidence. *NIPS '05: Proceedings of the 18th International Conference on Neural Information Processing Systems*, 1201–1208.

Schwartz, Odelia, Sejnowski, Terrence J., and Dayan, Peter. (2009). Perceptual organization in the tilt illusion. *Journal of Vision 9*(4).

Shea, Nicholas. (2018). *Representation in Cognitive Science*. Oxford: Oxford University Press.

Shea, Nicholas, Godfrey-Smith, Peter, and Cao, Rosa. (2017). Content in simple signalling systems. *The British Journal for the Philosophy of Science 69*(4), 1009–1035.

Sober, Elliott. (1976). Mental representations. *Synthese 33*(2–4), 101–148.

Solomon, Joshua A. and Morgan, Michael J.. (2009). Strong tilt illusions always reduce orientation acuity. *Vision Research 49*(8), 819–824.

Stalnaker, Robert. (1998). What might nonconceptual content be? *Philosophical Issues 9*, 339–352.

Strawson, P.F.. (1979). Perception and its objects. In G.F. Macdonald (Ed.), *Perception and Identity: Essays Presented to A.J. Ayer with his Replies to Them*, pp. 41–60. New York: Macmillan Press Ltd.

Szabó, Zoltán Gendler. (2012). The case for compositionality. In Wolfram Hinzen, Edouard Machery, and Markus Werning (Eds.), *Oxford Handbook of Compositionality*, pp. 64–80. New York: Oxford University Press.

Tacca, Michela C.. (2011). Commonalities between perception and cognition. *Frontiers in Psychology 2*, 1–10.

Tenenbaum, Joshua B., Kemp, Charles, Griffiths, Thomas L., and Goodman, Noah D.. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science 331*(6022), 1279–1285.

Tipper, Steven P. and Behrmann, Marlene. (1996). Object-centered not scene-based visual neglect. *Journal of Experimental Psychology: Human Perception and Performance 22*(5), 1261–1278.

Todorović, Dejan. (2010). Context effects in visual perception and their explanations. *Review of Psychology 17*(1), 17–32.

Treisman, Anne. (1986). Features and objects in visual processing. *Scientific American 255*(5), 106–115.

Treisman, Anne and Gormican, Stephen. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review 95*(1), 15–48.

Treisman, Anne and Paterson, Randolph. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance 10*(1), 12–31.

Treisman, Anne and Schmidt, Hilary. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology 14*, 107–141.

Watt, R.J.. (1990). Visual analysis and representation of spatial relations. *Mind & Language 5*(4), 267–288.

Wertheimer, Max. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed.), *A source book of Gestalt psychology*, pp. 71–88. London: Routledge & Kegan Paul.

Wijntjes, Maarten W.A. and Rosenholtz, Ruth. (2018). Context mitigates crowding: Peripheral object recognition in real-world images. *Cognition 180*, 158–164.

Xu, Haokui, Tang, Ning, Zhou, Jifan, Shen, Mowei, and Gao, Tao. (2017). Seeing “what” through “why”: Evidence from probing the causal structure of hierarchical motion. *Journal of Experimental Psychology: General 146*(6), 896–909.

Yuille, Alan and Kersten, Daniel. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences 10*(7), 301–308.

Zhu, Song Chun and Mumford, David. (2006). A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision 2*(4), 259–362.