Received: 1 November 2023 | Revised: 7 May 2024 | Accepted: 8 May 2024

DOI: 10.1002/wcs.1691

PERSPECTIVE

WIREs COGNITIVE SCIENCE WILEY

Compositionality in perception: A framework

Kevin J. Lande

Department of Philosophy and Centre for Vision Research, York University, Toronto, Canada

Correspondence
Kevin J. Lande, Department of Philosophy and Centre for Vision Research, York University, Toronto, Canada.
Email: lande@yorku.ca

Funding information
Social Sciences and Humanities Research Council of Canada; Canada First Research Excellence Fund

Edited by: Mohamad El Haj, Editor

Abstract

Perception involves the processing of content or information about the world. In what form is this content represented? I argue that perception is widely compositional. The perceptual system represents many stimulus features (including shape, orientation, and motion) in terms of combinations of other features (such as shape parts, slant and tilt, common and residual motion vectors). But compositionality can take a variety of forms. The ways in which perceptual representations compose are markedly different from the ways in which sentences or thoughts are thought to be composed. I suggest that the thesis that perception is compositional is not itself a concrete hypothesis with specific predictions; rather it affords a productive framework for developing and evaluating specific empirical hypotheses about the form and content of perceptual representations. The question is not just *whether* perception is compositional, but *how*. Answering this latter question can provide fundamental insights into perception.

This article is categorized under:

Philosophy > Representation
Philosophy > Foundations of Cognitive Science
Psychology > Perception and Psychophysics

KEYWORDS

compositionality, mental representation, perception, semantics, shape

1 | INTRODUCTION

One of the great challenges of modern science is to understand the immense creative powers of thought and language—our abilities to produce, understand, and communicate a seemingly infinite variety of spontaneous thoughts and plans, which can be arbitrarily unmoored from present concerns and circumstances. But impressive too is the immense responsiveness of perception: its expansive ability to reconstruct the immediate conditions in which it finds itself. Our perceptual systems are capable of representing endlessly varying and novel scenes, consisting of different arrangements of objects of different shapes, sizes, and textures, changing and interacting over time. We are capable of discriminating between these complex scenes while also being sensitive to elements that are common across them. The creativity of thought and language is thought to be powered by the capacity for *composition*—roughly, the ability to represent something (that Ida loves Una) by combining more elementary representations (of Ida, Una, and loving). Are the reconstructive capacities of perception also powered by compositionality?

Many philosophers have thought that it is the compositional machinery of language or thought that makes possible our abilities to construct representations of the world. Perhaps without thought, perception is a blooming, buzzing confusion of information (Dretske, 1981, p. 153). Perceptual representations might be simple "templates," or labels, each tuned to a specific stimulus and sharing no structure or parts with representations of other stimuli (Gross, 2002). They might be "blobs": points in a multi-dimensional psychological similarity space that lacks privileged axes (Lockhead, 1972). Or perceptual representations might be distributed over uninterpretable "units" in a network—they might have parts that do not in themselves represent anything (Chalmers, 1992; Smolensky, 1988). Without thought or language, perception might lack any kind of compositional structure. Another possibility is that insofar as there *is* compositionality in perception, it will be similar in form to the type of compositionality that resides in language and thought (McDowell, 1996). This might be the case, for example, if the role of compositionality in perception is to encode perceptual content in a format that is readable by cognition (Cavanagh, 2021; Quilty-Dunn et al., 2022).

I will argue that much of perception is compositional in its own right. The compositionality of perception is implicit in a wide variety of theories in perception science. David Marr once characterized perceptual representations as constituting a "formal scheme," comprising a set of elementary representations and "rules for putting them together" (1982, p. 21). Stephen Palmer argued that "perceptual representations are selectively organized data structures" (1977, p. 442). More recently, Jacob Feldman has emphasized that "compositional representations are [...] useful and effective mechanisms for representing complex probabilistic phenomena in the environment" (2023, p. 3). But compositionality can come in a variety of forms. I argue that compositionality in perception is quite unlike the compositionality that is thought to characterize language and thought. The claim that perception is compositional is best understood not as a concrete hypothesis that makes specific predictions, but rather, like Bayesianism (Griffiths et al., 2012; Knill et al., 1996) or neuroconnectionism (Doerig et al., 2023), as a productive theoretical framework for posing and answering empirical questions about perceptual representations. The question is not simply *whether* perception is compositional, but *how*. Answering this latter question can provide fundamental insights into the nature of perception and its place within the mind.

2 | WHAT IS COMPOSITIONALITY?

When one asks about the "composition" of a photograph, a novel, or a mathematical function, one is asking how one thing depends for certain of its properties on the way it is combined from some separate parts and certain of their properties. The relevant notions of "thing," "part," and "combination" can vary depending on the subject matter (e.g., a rock song and the temporal arrangement of its musical phrases vs. a sample of rock and its physical aggregation of constituent minerals), as can the relevant properties (one can ask about a sculpture's aesthetic composition or its chemical composition). *Semantic compositionality* is specific to representations—roughly, things (sentences, maps, pictures) or psychological states (perceptions, memories, thoughts) that matter in large part because of what they are about.

A newspaper headline is a representation. Besides being written in letters that have certain shapes and sizes, it conveys content about some world event. Your visual state is also a representation. It does not merely involve a set of neurons firing; critically, it also carries content about the shapes, orientations, and motions of the things around you. Some representations, such as the newspaper headline, have other representations (words) as parts. Perceptual representations, I will argue, also have parts. Semantic compositionality concerns the relationship between what a whole representation is about (its "content") and what its parts are about. As I emphasize later, this relationship does not directly concern how a representation is formed or how it is decoded downstream, although these are certainly not independent of the compositionality of the representation. Semantic compositionality is first and foremost a property of *representations*, not of *processes*.

The concept of a *semantically compositional representation* has its roots in philosophy of language, logic, and linguistics, and is commonly taken to have originated with the work of Gottlob Frege (b. 1848, d. 1925). It is standardly defined as follows (Janssen, 2011; Partee, 2004):

> *Compositional representation*: A representation is *compositional* if and only if its content depends wholly on the contents of its constituent parts and the way those constituents are structurally combined.

For example, the phrase "red balloon" is a compositional expression. We can characterize the content, or meaning, of the phrase by noting that it is true of anything that is both red and a balloon. It applies to just these things because

(i) "red" is an adjective that is true of anything that is red and "balloon" is a noun that refers to the set of things that are balloons, (ii) "red balloon" is a combination of the adjective "red" and the noun "balloon," and (iii) in the standard case, a combination of an adjective, A, and a noun, N, is true of anything that is a member of the set designated by N and of which A is true (Heim & Kratzer, 1998, pp. 65–67).

More generally, if a representation is compositional, then one can exhaustively characterize its content (what it is about) by answering what I will call the "Three Cs":

The Three Cs:

1. What are the elementary, or primitive, *constituents* of the representation?
2. In what way are the constituents structurally *combined*?
3. How does the *content* of a representation depend on the contents of its constituents, given the manner in which they are combined?

We can also pose the Three Cs for a system as a whole, which is tantamount to asking about the system's "lexicon," its "syntax," and its "compositional semantics," respectively. A set of representations is a *compositional system* insofar as each complex representation in the system is compositional and, moreover, there are common principles—common answers to the Three Cs—that determine the contents of multiple representations in the system. For example, "white rabbit," "square hat," and "litigious pony" are all governed by the A-N principle.

Content, *constituent part*, and *structural combination* all are theoretical concepts. While theorists differ in how they understand and even label these concepts, most agree on some core points. First, an essential feature of a representation is that it has the function of being about something. The "content" of a representation—as with the content of a newspaper headline—determines what it is about and what would be required for the representation to be a correct characterization of the world (Burge, 2014). The content of the sentence "Ida loves Una" is about Ida loving Una, and the sentence is true just in case Ida *does* love Una. The content of the monkey's current visual state is about (among other things) the estimated orientation of the target Gabor patch, and that visual state is accurate insofar as the target Gabor *has* that orientation.

Second, much as physical laws determine what molecules are structurally possible given a set of atomic elements, so structural constraints or "combination rules" (Feldman, 2023) are psychological laws that determine what complex representations are structurally possible given a set of elementary representations. "Ida loves Una" is a possible sentence in my idiolect because it has the right kind of constituents, combined in the right kind of way. "Ida Una balloon luft" does not, and so is not a possible sentence in my idiolect (though it is physically possible for me to pronounce and perceive the physical string). A theory of representational structure (a theory of syntax or grammar) aims to explain what representations are possible in a system—and therefore available for computing or inferring with—in terms of the ways representations can, cannot, or must combine to be constituents of a composite representation. A theory of representational content (a theory of semantics) aims to explain what those representations are about (Heim & Kratzer, 1998; Larson & Segal, 1995). In a compositional system, the same principles that delimit which representations are possible also delimit what those representations are about.

That a system is compositional is not a concrete hypothesis that makes specific empirical predictions. It is a theoretical framework for formulating such hypotheses. To hypothesize that a system is compositional is simply to say that there is some way of filling in the Three Cs that explains the space of what can be represented in that system. Specific predictions are obtained only when one hypothesizes specific values for those three parameters.

Still, it is an empirical question whether compositionality offers a feasible framework for studying a given system. Filling in the Three Cs is no trivial task. As the formal semanticist Barbara Partee (2014) recounts, prior to the research programs championed by Noam Chomsky and Richard Montague, many doubted that meaning in natural language could be explained in terms of the Three Cs (see also Davidson, 1967). Many still doubt it (Pagin & Westerståhl, 2010). Beyond language, the claim that thinking and planning are compositional (Fodor, 1975; Quilty-Dunn et al., 2022) has proven to be controversial. In cognitive science, discussions of compositionality have largely focused on language and thought, occasionally touching on the possibility that "high-level perception"—perceptual capacities such as object recognition that sit at the interface with language and thought—might involve compositional representations. I will argue that in fact a great deal of perception science operates under the working assumption that perceptual representations, at virtually all levels of processing, are compositional. This assumption of compositionality is often implicit and unacknowledged. We can better evaluate and design studies of perceptual representations by making explicit the commitment to compositionality as a working framework. I focus here on vision.

3 | COMPOSITIONALITY IN VISION

Much of vision is compositional. Vision science contains a host of hypotheses about how one type of stimulus feature is represented or coded "in terms of" other features. These constitute hypotheses about how one representation is composed of others (Figure 1). Here are some examples:

- A representation of a surface's orientation in depth is composed from representations of its *slant* and *tilt* (Nguyenkim & DeAngelis, 2003; Stevens, 1983).
- A representation of an element's motion is composed from a representation of the common motion that it shares with other elements with which it is grouped and a representation of the element's residual motion (Gershman et al., 2016; Johansson, 1973).
- A representation of visual texture is composed from representations of summary statistics of orientation and other features (Balas et al., 2009; Portilla & Simoncelli, 2000).
- A representation of an object's orientation relative to an extrinsic object (e.g., the orientation of my pen on the notepad) is composed from representations of "(a) a correspondence between object axes and external axes, (b) the tilt of the object axes relative to the external axes, and (c) the relationship between the polarity of the object axes and the polarity of the external axes" (McCloskey et al., 2006, p. 685).
- A representation of an object in vision and visual memory (an "object file") is composed from representations of features—including its orientation, motion, texture, and so on—perhaps with a separate "index" for the object that is invariant across feature changes (Green & Quilty-Dunn, 2017; Kahneman et al., 1992; Pylyshyn, 1989; Treisman, 1986).
- A representation of a scene is composed from representations of objects and their relations (Biederman et al., 1982; Hafri & Firestone, 2021; Võ, 2021; Zhu & Mumford, 2006), or from representations of summary statistics of features such as the mean depths of surfaces within the scene (Greene & Oliva, 2009).

Even if, as is very likely, none of these hypotheses is exactly right, that they all presuppose compositionality suggests that this is a productive framework for honing in on the truth. In the rest of this section, I look at how compositionality

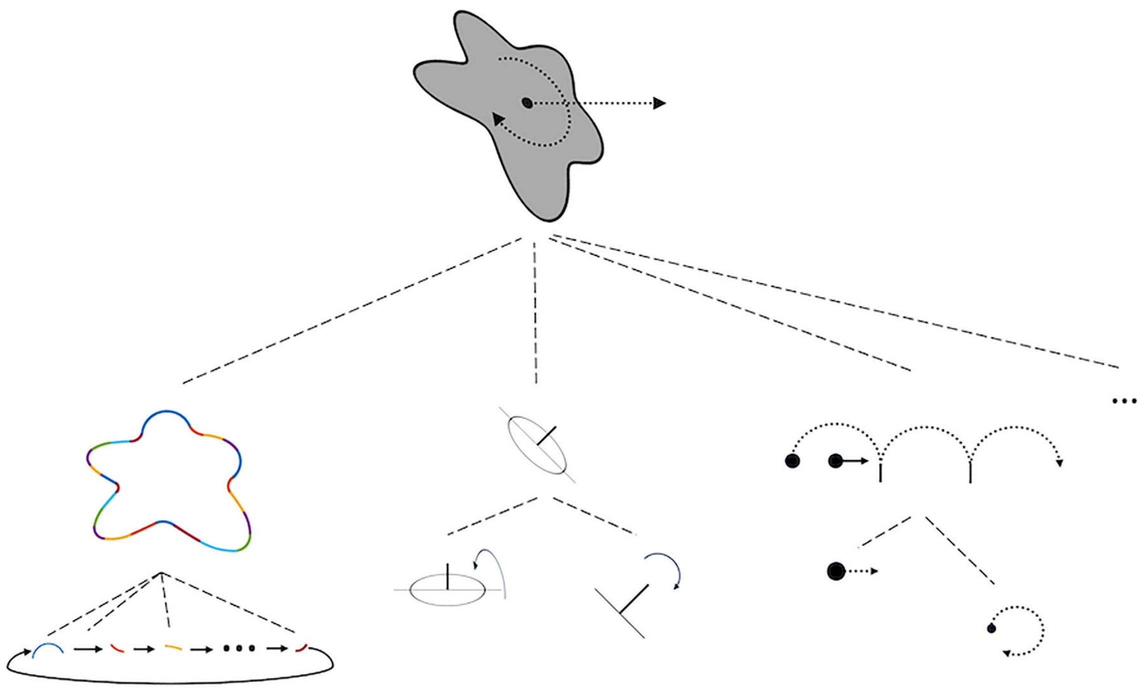

FIGURE 1 Vision is compositional. A spinning object moving through the scene is represented in terms of (among other things) its shape, 3D orientation, and motion, which themselves may be represented, respectively, in terms of configurations of shape parts, slant and tilt, and common and residual motion vectors.

structures the way many researchers have investigated one of the fundamental problems in vision science: how humans and primates represent shape.

3.1 | Compositionality in shape perception

Standard approaches to shape perception assume compositionality as a working principle. Here, I survey some different hypotheses concerning the Three Cs of visual shape representation (Figure 2; Table 1). I do not mean to imply that these accounts are all equally likely to be true, only that they all fall under the broad church of compositionality. Should one hypothesis be more well-supported than the others, it would reflect something about *how* shape representations are compositional and not just *that* they are. For the sake both of simplicity and of looking beyond high-level perception, I focus on accounts of how the two-dimensional outlines of things are represented in "mid-level" vision, which is concerned with representing cohesive figures and groups within a scene (Elder, 2018; Todd & Petrov, 2022; see also Lande, 2023).

"Part-based" accounts of shape perception hold that the perceptual representation of a shape is composed from constituent representations of parts of that shape and that structural relationships between constituents reflect configural relationships between the corresponding parts. For example, according to "contour-based" part theories (Figure 2a–c; Table 1a–c), the relevant shape-parts are segments along the object's outline, or bounding contour. Some models take the constituents to be representations of local oriented edges (Elder & Goldberg, 2002; Geisler et al., 2001). Most take the constituents to be representations of curved segments, whether just the convex segments or bumps

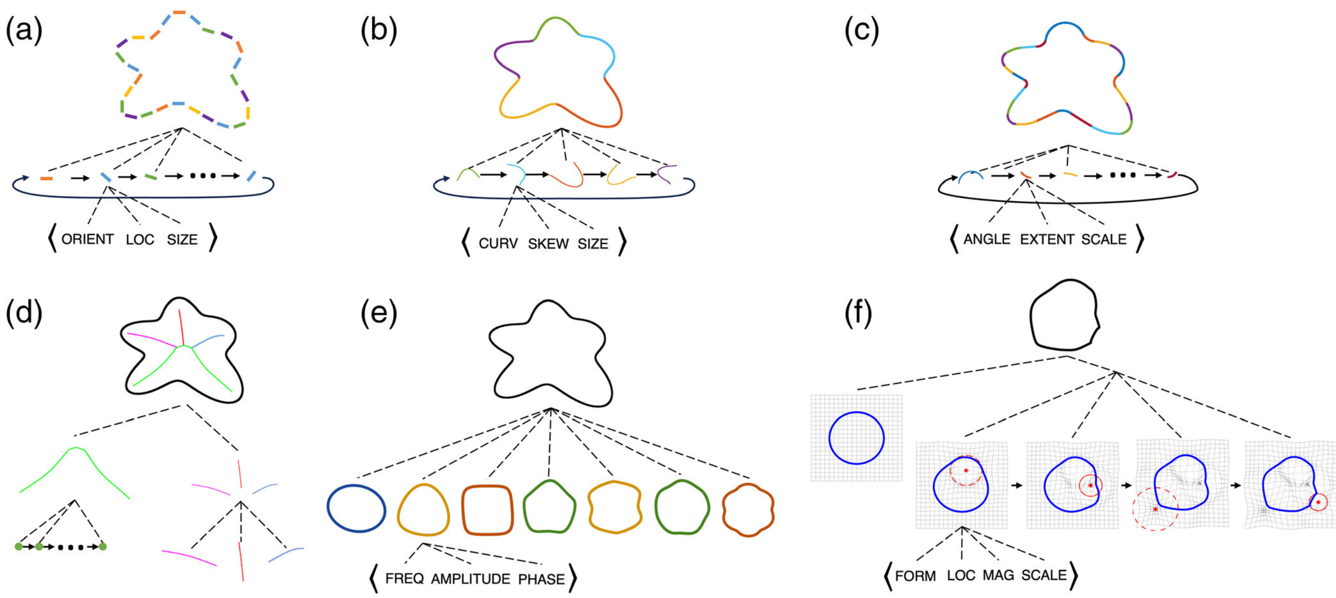

FIGURE 2 Schematic illustrations of different compositional models of two-dimensional shape representations. Dashed lines indicate constituency. Solid lines indicate structural relationships between co-constituents—for example, that one representation precedes another within an ordered sequence. Some constituents decompose further into representations of orientation, location, size, and so forth. See Table 1 for corresponding combinatorial constraints and their semantic import. (a) An edge code, in which shape representations are composed from an ordered sequence of representations of oriented line segments (Elder & Goldberg, 2002; Geisler & Super, 2000). (b) A minima or convexity code, in which shapes are represented in terms of convex segments that are segmented at curvature minima, that is, dents or points of extreme concavity (Cohen & Singh, 2007; Hoffman & Richards, 1984; Richards & Hoffman, 1985; see also Bertamini & Wagemans, 2013; Haushofer et al., 2008; Liu et al., 2010; Schmidtmann et al., 2015). (c) A constant curvature code, in which shapes are represented in terms of segments of constant curvature (Baker et al., 2020; Kellman et al., 2013; Pasupathy & Connor, 2001). (d) A medial axis code, in which a shape is represented in terms of branching axes of symmetry (Blum & Nagel, 1978; Feldman et al., 2013; Feldman & Singh, 2006; Green, 2023; Kimia, 2003). (e) A Fourier descriptor code, which represents shape in terms of the amplitude of the shape's curvature at different frequencies (Cortese & Dyre, 1996; see Elder, 2018). (f) A formlet scheme in which shapes are coded in terms of sequences of warping transformations applied to an embryonic template shape (Elder et al., 2013; see also Leyton, 1988, 1992) (f is adapted from Elder et al., 2013, p. 5 with kind permission from Elsevier).

TABLE 1 Compositional schemes for representing shape differ in (1) what they take to be the constituents of shape representations, (2) how those constituents can be combined (including the type of structural relations they can stand in when combined—their "form"—and the constraints they must satisfy to stand in those relations), and (3) how the content of the whole shape representation is derived from the contents of its constituents and their manner of combination.

System	(1) Constituents	(2) Combination	(3) Content
(a) Edge code	Representations of oriented edges, composed from representations of • Orientation • Location • Size	• Form: Ordered sequence (or cycle) • Constraint: Good continuation	A shape that has each represented edge as a part, such that adjacent constituents represent neighboring edges
(b) Minima/convexity code	Representations of convex contour segments, bounded at extreme points of concavity, composed from representations of, for example, • Curvature • Skew • Size	• Form: Ordered sequence (or cycle) • Constraint: Good continuation	A shape that has each represented segment as a part, such that adjacent constituents represent neighboring convex parts
(c) Constant curvature code	Representations of constant curvature segments, composed from representations of • Turning angle • Extent • Scale	• Form: Ordered sequence (or cycle) • Constraint: Good continuation	A shape that has each represented segment as a part, such that adjacent constituents represent neighboring parts
(d) Medial axis code	Representations of local axes of symmetry, composed from, for example, representations of points along the axis	• Form: Hierarchical tree in which each constituent may be adjoined by a set of secondary constituents that represent protruding axes • Constraints: ○ Smoothness ○ Simplicity	A shape each part of which is roughly symmetrical about a represented axis and where the arrangement of axes reflects the hierarchical organization of constituents
(e) Fourier code	Representations of amplitude and phase at particular frequencies	• Form: Ordered sequence • Constraint: Constituents are ordered from coarsest to finest frequencies	A shape that results from the superposition of the different frequency components
(f) Formlet code	• Embryos: representations of basic template shapes • Formlets: representations of warping transformations with a specific location, magnitude, and scale	• Form: Ordered sequence • Constraint: Initial constituent is an embryo shape; subsequent constituents are formlets	A shape that results from the sequential application of warping transformations to the embryonic shape

Note: See Figure 2 for illustrations of these different schemes.

(Hoffman & Richards, 1984; Richards & Hoffman, 1985; De Winter & Wagemans, 2006; Schmidtmann et al., 2015; Figure 2b), or segments of constant curvature that approximate parts of the object's contour (Baker et al., 2020; Kellman et al., 2013; Pasupathy & Connor, 2001; Figure 2c). These part-representations themselves decompose further into representations of features such as curvature, size, and so on.

Many contour-based schemes assume that shape representations have the form of ordered sequences or strings (in the case of closed figures, cycles) in which one part-representation is structurally adjacent to another (e.g., Elder & Goldberg, 2002; Figure 2a–c; Table 1a–c). Standardly, whether two representations can combine as adjacent constituents is a function of whether they represent segments as having positions and orientations that are relatively collinear and nearby (Boucart et al., 1994; Elder, 2018; Field et al., 1993; Kellman & Fuchser, 2023; Khuu et al., 2016). This combination rule reflects the traditional Gestalt principle of "good continuation" (Lande, 2021). Given this rule, if two part-representations are structurally adjacent, this signifies that the represented parts are topologically adjacent and more or less well-aligned. So, the constituents contribute content about the parts of the shape (their curvature, size, and so on), and the structural relationship between the constituents (the manner in which they are combined) contributes content about the relative positions and alignment of those parts (Lande, 2023).

Rather than representing an object's shape in terms of segments of its boundary, "region-based" representations code shape in terms of features of the bounded regions, such as their axes of symmetry. For example, your palm and your fingers are each roughly symmetric about their own axes. A "medial axis representation" codes the shape of your whole hand in terms of these different local axes of symmetry (Blum, 1973; Feldman et al., 2013; Feldman & Singh, 2006; Kimia, 2003; Figure 2d; Table 1d). As Green (2023) and Hafri et al. (2023) emphasize, medial axis representations are compositional. A medial axis representation of a shape has a tree-like structure. The main body of a shape (e.g., the palm of a hand) is represented in terms of a main axis. The representation of this main axis is adjoined by a set of constituents that represent the axes of protruding parts (e.g., the fingers). Each of these may in turn be adjoined by sets of constituents that represent further offshoots. Each axis is itself coded, for example, as a sequence of points (Feldman et al., 2013, p. 57). A smoothness constraint entails that an axis representation is more likely to decompose into a collinear set of points (related by shallower turning angles) than a ragged or twisty set of points, and a simplicity constraint entails that there is a fixed cost to adjoining axes to a shape representation. While the constituents of a medial axis representation carry content about the local axes around which parts of the shape are symmetrical, the specific way the axis representations are combined contributes content about how the represented shape-parts are related, for example, which regions protrude from which others. So, as Green writes, "The syntactic structure of a [medial axis] representation mirrors [...] relations among parts of the shape it represents" (2023, pp. 10–11).

As Green notes, compositional theories of shape representation are not all part-based (2023, pp. 40–41). For example, work in computer vision as well as some studies in biological vision appeal to Fourier descriptors (Figure 2e; Table 1e), which are a *global* contour-based scheme for representing shapes (Cortese & Dyre, 1996; Elder, 2018, pp. 436–437). Rather than representing a shape piecewise in terms of features of its distinct segments, a Fourier code consists of constituents that represent the phase and amplitude of a whole outline's curvature at different frequencies, from coarsest to finest. The represented shape is the superposition of all these frequency components.

In contrast to the above "static" models of shape perception, "process-historical" schemes represent shape in terms of the process, or series of transformations, from which the shape could be generated (Leyton, 1988, 1992; Spröte et al., 2016). For example, in a "formlet" scheme (Elder et al., 2013; Figure 2f; Table 1f), a shape representation has the form of an ordered sequence, where an initial constituent represents an embryonic shape such as an ellipse and other constituents represent localized warping transformations or distortions to the shape. On some process-historical schemes, each spatial "part" of a shape corresponds to a specific sequence of transformations. These schemes are "part-based" only in a loose sense, for the shape's parts are neither represented as contour segments nor as regions—rather, "A part is a process!" (Leyton, 1992, p. 31).

This brief survey is not meant to suggest that each of the schemes mentioned above is equally plausible or well-supported. But it is striking that they all presuppose the compositionality of shape representations. Where they differ is in what they take to be the constituents, combination rules, and content rules of those representations—that is, how they answer the Three Cs. Compositionality offers a framework for articulating these schemes and their differences. It also provides a framework for evaluating them.

3.2 | Diagnosing structure

Though a great deal of experimental ingenuity may be required to tease apart alternative answers to the Three Cs, there is tacit consensus about what kinds of tests would in principle help to evaluate those alternatives. The experimental logic for tests of constituency and combinability (I set aside tests of content) rest on a few defeasible principles (Lande, 2021; Schwartz & Sanchez Giraldo, 2017). Making these principles explicit can encourage more systematic research on the Three Cs of shape perception, not to mention the perception of other features such as texture and motion.

1. The occurrence of a representation requires the occurrence of its constituents.

Without the parts, you cannot have the whole (Fodor & McLaughlin, 1990, p. 198; Schwartz & Sanchez Giraldo, 2017, p. 6). For example, if a shape representation is "made out of" certain edge representations (Figure 2a), then if those edge representations are unavailable the shape representation that they compose will be unavailable too.

Introducing noise to a shape's edge elements should therefore diminish the sensitivity with which a perceiver detects and discriminates the overall shape. So, the hypothesis that shape is coded in terms of oriented edges predicts that substantial perturbations to local edge information should reduce sensitivity to the overall shape. This prediction is disconfirmed by experiments showing that shape discrimination can be quite resilient to variations in local edge information so long as overall curvature information is preserved (Baker & Kellman, 2018; Boucart et al., 1994).

By contrast, curvature-based codes predict that masking curvature information would substantially impact shape representation. If shapes are represented in terms of curvature, then noisy curvature representations would mean noisy shape representations. This prediction has been borne out with multiple stimuli and paradigms (De Winter & Wagemans, 2006; Gheorghiu & Kingdom, 2007; Habak et al., 2004; Schmidtmann et al., 2015), suggesting that even if local edge orientations can be *cues* to shape, at some processing stage shape is *coded in terms of* curvature.

Suppose that the shape of a hand is coded in terms of the curvature of the hand's outline. Does each finger get its own constituent in the representation of the hand? What about different parts of each finger? To answer this sort of question, we turn to a second principle:

2. Distinct constituents are conditionally independent of each other, given that they belong to the same representation.

Put another way, a constituent can be preserved ("reused") through variation or substitution of the other constituents with which it is combined. So, to test whether different features or parts of a stimulus receive separate constituents in the stimulus' representation, one can look at how variability in the representation of one part or feature corresponds to variability in the representation of another (Garner, 1974; Schwartz & Sanchez Giraldo, 2017, pp. 6–10). If subjects represent the thumb in the same way no matter the shape, size, or presence of other fingers, then this suggests that the thumb has its own, dedicated unit in the representation of the hand. For example, selectively masking or adapting subjects to different curved segments of a contour, or to different amplitudes of curvature, reveals both that distinct frequencies and amplitudes of curvature are encoded separately, but also that these properties are encoded separately for distinct parts (Gheorghiu & Kingdom, 2007; Habak et al., 2004). Likewise, Cohen and Singh (2007) found that subjects were systematically more sensitive to variation in bumps of a shape's contour (i.e., segments bounded at points of high concavity) than to variation in other segments along the contour. On the assumption that responses that are based on one cohesive constituent or "unit" in a representation will be more sensitive than responses based on multiple independent constituents (Palmer, 1977), these data speak to a part-based scheme in which distinct segments of a shape receive distinct constituents in the shape's representation rather than a global scheme such as a Fourier code.

To be clear, one shouldn't expect perfect or absolute independence between separate constituents. One of the strengths of the compositional framework is that one can express *interdependencies*, or "co-determination," among constituents (Lande, 2021; Schwartz & Sanchez Giraldo, 2017; Zhu & Mumford, 2006). The rule of good continuation reflects the fact that the orientation of one segment along a contour is not completely independent of the orientation of a neighboring segment—the latter is unlikely to be perpendicular to the former, for example. As a result, if the visual system is biased to combine representations of two shape segments into a representation of a cohesive shape, then the orientations of the segments might be represented as more collinear than if the visual system were to treat the segments as separate (Keemink & van Rossum, 2016; Schwartz et al., 2009). Nevertheless, there are multiple ways in which one segment could be a good continuation of another and, within that range of alternatives, the represented orientations of the two segments are approximately independent of each other.

There has been somewhat more focus in vision science on employing the above two principles to identify the separate constituents of a representation (the first C) than on specifying the constraints on how those constituents can combine (the second C). An important principle in investigating combinatorial constraints is:

3. Combinatorial constraints determine which representations are possible in the system as a function of how those constituents are related.

The posited combinatorial constraints should not "under-generate" representations: they should not count as impossible representations that seem to be psychologically possible. Nor should they "over-generate," counting as possible certain representations that seem to be psychologically impossible. For example, a prediction of the hypothesis that good continuation is a constraint on combining contour representations is that without any good continuation at any spatial scale, you won't be able to represent an integrated contour (Field et al., 1993). This implies that, as Brian Keane (2018) writes, "for certain element layouts, there is no amount of learning or top-down strategy that will get the elements to link together in a spatially precise way" (p. 13). Indeed, while task set can influence which contour segments in a crowded scene are more likely to be integrated on a given occasion, the represented contours always fall near or within the range of good continuation (McManus et al., 2011). This is evidence that good continuation is a combinatorial constraint on contour representations—that it dictates whether a contour representation is possible as a function of what its constituents are and how they are related (Lande, 2021).

It is likely that perception employs multiple representational schemes to encode a given property such as shape (Todd & Petrov, 2022). Systematically applying the above three principles can help not just to confirm or disconfirm competing hypotheses about the Three Cs, but also to reveal different families of compositional representations, each of which is explained in terms of its own proprietary set of primitives and combination rules.

Compositionality provides an effective and tractable framework for formulating and testing hypotheses about the way stimuli are encoded in perception. I have given the study of shape perception as an example, but I also alluded above to examples in the visual perception of texture, motion, orientation, and so on. A great deal of vision science works within the framework of compositionality, and this work has been, and will likely continue to be, fruitful. This is good reason to think that much of vision is in fact compositional. Still, the framework and its experimental logic often are only implicit in perception research. Making these explicit can encourage more systematic and consistent development, testing, and comparison of models, allowing us to better understand *how* vision is compositional.

4 | DOES VISION HAVE A LANGUAGE?

Some hold that insofar as perceptual representations are compositional, they have a similar form as representations in cognition and language. It has been suggested that perceptual representations have a "grammar" (Gregory, 1970) or "logic" (Rock, 1985), that they are like "descriptions" (Hummel, 2013) with "logical predicates" (Feldman, 1997), and that there is a "language of vision" (Cavanagh, 2021). On a modest interpretation, these claims simply imply that perceptual representations, much like language (according to the standard view in linguistics), are compositional. But, to reiterate a distinction emphasized by Elisabeth Camp (2007), the claim that perception and language are both compositional does not imply the much stronger conclusion that perception and language are both compositional *in the same ways*. By and large, the primitives and combination rules of perception are not very language-like.

Perception lacks much of the expressive power of language, logic, and conceptual thought. Consider, first, the "lexicon" of constituents available in perception. There is no evidence for explicit representations of logical relations (*not*, *or*, *if … then …*) in perception (Block, 2023, chapter 3). Moreover, many categories of words and concepts are open class—you can endlessly expand your stock of nouns and verbs, thing-concepts and property-concepts. Correspondingly, there are few constraints on what a noun (or thing-concept) can represent: a person, a transfinite cardinal, or a weather event. By contrast, the primitive dimensions along which we can perceptually represent things (orientation, size, and so on) are arguably quite fixed (Green, 2020). Perception plausibly lacks the open-class types of primitives that characterize the elements of natural language and conceptual thought.

The expressive power of perception is limited not just by the types of primitives available in perception but also by the available combination rules (Lande, 2021, 2023). Nouns with quite different meanings can be interchanged within a sentence without threatening its grammaticality ("Ida loves [Una / aleph-null / the weather]"). In fact, this pattern of intersubstitutability defines a noun as a syntactic type. There is much less freedom for substitution between perceptual representations. For example, Green and Quilty-Dunn (2017) argue that visual object files have "feature-specific slots." Roughly speaking, an object file can be composed from *one* shape representation, *one* orientation representation, *one* size representation, and so on. Shapes can be swapped for shapes and orientations for orientations; but shapes cannot enter into the orientation slot or vice versa (see also Ashby, 2020). In language and thought, though "rectangular" and "upright" are semantically different, they have the same syntactic role. In perception, "rectangular" and "upright" differ both in their semantic and syntactic roles. The combination rule of good continuation provides another example. In language, words can combine only if they agree in relatively abstract features such as person, number, and case. In vision, contour representations can combine only if they "agree" in spatial configuration, by representing contours with sufficiently aligned positions and orientations. The compositional roles of perceptual representations are much finer-grained and domain-specific than those that of nouns and adjectives, say.

The domain-specific form of the perceptual "grammar" endows perceptual representations with an epistemic character that most linguistic expressions lack. The grammar of language permits us to express unexpected and even contradictory claims. This incurs the risk of saying and thinking highly inaccurate things, but it also enables us to state and argue for deep and surprising insights—that the apple and the moon are governed by the same laws of motion, or that there are infinitely many prime numbers. Perception is notably more conservative. It is unlikely that the motion of the moon across the sky and the motion of the apple falling to the ground would ever be grouped together visually, while the sorts of contours that can be represented given the combinatorial constraint of good continuation are just the sorts of contours that tend to arise in natural scenes (Elder & Goldberg, 2002; Geisler et al., 2001). In language, well-formedness guarantees neither plausibility nor even meaningfulness ("Colourless green tomatoes sleep furiously"). In perception, the combinability of representations tends to track more closely the likelihood that the compound representation would be accurate.

It is often said that the symbols of language have an "arbitrary" connection to their subject matter and can be understood "according to certain purely syntactic rules" that make no reference to what they mean (Newell & Simon, 1976, p. 117; see also Fodor, 1983). Rather than signifying round bags of air, the word "balloon" could have meant *transfinite cardinal*, and this difference in meaning would not entail any difference in the syntactic form or logical role of the phrase "red balloon." The above comments suggest that the grammar of perception, such as it is, has a much less arbitrary connection to its subject matter than is often claimed for the grammar of a language. An adequate account of the specific rules by which perceptual representations can be put together may well require understanding what features of the world those representations are about and the ways in which those features tend to mix together in the perceiver's environment (for different aspects of this point, see Burge, 2010, pp. 19-25; Neander, 2017, Chapter 8).

One motivation for thinking that compositionality in perception would be like compositionality in cognition stems from the idea that compositional representations are a way of packaging perceptual content into a format that is readable by language and thought (Cavanagh, 2021; Quilty-Dunn et al., 2022). But we have seen that compositionality is posited throughout vision, including in the mid-level processing of outline shapes, and not just at the high-level interface with cognition. The constituents of orientation, texture, motion, and shape representations may be far removed from the kinds of features we tend to talk and think about when describing our visual world. Moreover, the compositional principles of perception seem to subserve a central function of perception itself: to accurately represent the distal causes of sensory stimulations (Graham, 2014; Knill et al., 1996; von Helmholtz, 1925). The perceptual code ensures that the hypothesis space of perception is well-fitted to its ecological niche, containing representations that are plausible given the laws and regularities of the perceiver's normal environment and that have constituents that glom onto meaningfully separate sources of variability in the world. By contrast, while constraints of plausibility may guide conceptual inference, they do not dictate which thoughts are thinkable. The compositional principles governing perception may in some cases serve to facilitate the interface with language and thought; but in many cases they may be better understood as subserving endogenous functions of perception itself.

5 | WHAT COMPOSITIONALITY IS NOT

I now want to highlight some claims that are *not* consequences of semantic compositionality. In perceptual psychology and computer vision, the term "compositional representation" is often taken as synonymous with "part-based representation," whereby something is represented in terms of its parts and their relationships. But semantically compositional representations need not strictly be part-based. Visual representations of texture are semantically compositional, but according to one account their constituents represent the statistics of a texture rather than its individual parts (Balas et al., 2009). Global models of shape perception, such as Fourier descriptors, represent shapes in terms of features of the whole shape. Language, which is widely thought to be compositional, is rarely part-based. The phrase, "The 16th President of the United States," is a compositional representation that is about Abraham Lincoln; but the constituent, "the United States," does not represent a literal part of Abraham Lincoln (Block, 1983). If some perceptual representations are in fact part-based, then that is a substantive fact about the way they are compositional.

Semantic compositionality is a feature of representations, not processes; it is consistent with different ways in which those representations might be formed and used (Lande, 2021). Talk of "combination" and "composition" might encourage a picture of processing, according to which the constituents are formed first and then "put together" (Treisman & Gelade, 1980). But semantic compositionality does not entail an assembly line. By analogy, that the human body is *composed of* limbs and other body parts does not mean that it was *assembled from* those parts, as if the parts were first individually fabricated and then put joined together on a factory line. Compositional representations may be susceptible to configural effects, top-down influences, and recurrent processing. Some compositional theories of vision posit an "analysis-by-synthesis" architecture in which compositional representations are refined through the convergence of bottom-up cues and prior expectations and constraints (Yuille & Kersten, 2006). More generally, when confronted with a particular stimulus, multiple structural representations may be activated and compete with each other for priority (e.g., Dakin & Baruch, 2009; Li, 1998). Configural factors such as what other stimulus elements are perceived (Kovács & Julesz, 1993; Palmer, 1977) as well as "top-down" factors such as familiarity or expectation might bias one structured representation over the other (McManus et al., 2011; Sassi et al., 2014). Similar points apply to other compositional systems, such as language. How one parses one part of a sentence can be influenced by the rest of the sentence as well as by cognitive context (Bever & Poeppel, 2010). In a compositional system, contextual information can *bias* which possible combination wins out without violating compositionality.

Moreover, while we can expect attention and memory to be able to operate on any of the constituents of a compositional representation, there is no requirement that they must always make use of all of the structure in a representation (see Barenholtz & Tarr, 2006). Recognition may succeed (or fail) before each and every constituent of a visual representation has been matched to representations in memory. For success on some perceptual tasks, such as visual search or classification, it might suffice for downstream processes to decode the presence of a single task-relevant constituent in a complex representation while ignoring task-irrelevant constituents, or to decode the structural form of a representation (e.g., that it is a cyclical ordering of constituents, signaling a closed contour) abstracting from what its specific constituents are. To probe what the constituents of a representation are, one would ideally use a variety of tasks or employ paradigms that minimize the fingerprints of specific downstream processes.

Finally, it is helpful to distinguish *psychological* (or *cognitive*) representation from *neural* representation. Psychological representations are typically posited as elements of models that explain psychophysical patterns in how individuals experience and behaviorally respond to stimuli (Burge, 2014; Palmer, 1978). Neural representations are patterns of neural activity that are investigated through various neuroscientific probes, such as neuroimaging and extracellular recordings (Cao, 2022). A central task of cognitive neuroscience is to test "linking hypotheses" that relate the sorts of representations posited in psychological models to the sorts of representations probed in neuroscience (Baker et al., 2022; Shea, 2018; Teller, 1984). Instead of presupposing that a compositional psychological representation should look a certain way in the brain, these linking hypotheses are empirical conjectures that are subject to theoretical evaluation and experimental dis(confirmation). In the case of compositionality, some linking hypotheses are straightforward—for example, that distinct psychological constituents correspond to distinct cells or neural ensembles (Vaziri et al., 2009), or that one compositional model of stimuli can jointly predict both neural responses and behavioral responses (Arun, 2022). But there is no guarantee that the link between psychological compositionality and neural function will be so neat.

Likewise, while many deep neural network (DNN) models of vision seem to eschew compositionality, DNNs are not inherently non-compositional. Ongoing work seeks to identify what properties of these networks and their training regimes might give rise to compositional representations as well as the advantages and challenges of incorporating structured representations into those networks (Lake et al., 2016; Peters & Kriegeskorte, 2021). Since it is often not transparent whether a DNN has come to learn representations that are compositionally structured, diagnostics of the sort canvassed Section 3.2 can be applied to networks, possibly providing evidence of representations with compositional structure like that diagnosed in biological vision (Hupkes et al., 2020; Lepori et al., 2023; Zerroug et al., 2022). In any case, uncertainty about either biological or computational linking hypotheses does not invalidate the existing psychophysical methods for evaluating hypotheses about the structures of psychological representations. Indeed, such methods are critical for validating candidate linking hypotheses.

6 | CONCLUSION

Perception involves encoding content or information about the world. In what form is this content represented? I have argued that perception is compositional: in many cases, the content of a perceptual representation depends wholly on the contents of its constituents and the manner in which they are combined. Vision science contains manifold models of how shape, orientation, motion, and other aspects of a scene are "coded in terms of" various features. These models explain the space of possible representations in a system by answering the Three Cs: what are the constituents, how can they combine, and what is the semantic import of combining them that way? Compositionality provides a productive framework for understanding perceptual representation.

There are no guarantees, to be sure, I have not been able to cover non-visual perceptual modalities or cases of multi-modal perception. Perhaps some perceptual representations are best thought of as atomic templates

(Gross, 2002), holistic "blobs" (Lockhead, 1972; cf. Kemler Nelson, 1993; Jones & Goldstone, 2013), or ensembles distributed over "parts" that in themselves "carry no semantic burden" (Chalmers, 1992; Smolensky, 1988). It is an empirical question *how many* perceptual capacities are compositional and, for any given capacity, *how* it is compositional.

I close by highlighting three principal conclusions. First, the view that there is no compositionality in perception does not sit well with much of contemporary perception science. Second, compositional representations in perception are markedly different from those in cognition. Structured representations in perception do not simply sit at the interface with cognition, serving to relay messages in a format readable to it. The form in which perceptual content is represented supports perception's own proprietary ends. Finally, I have stressed that the thesis that perception is compositional does not constitute a concrete hypothesis with specific empirical predictions. It instead offers a productive theoretical framework for posing and answering empirical questions about the nature of representations in a system. Fundamental insights arise within that framework when one asks not just *whether* perception is compositional, but *how*.

AUTHOR CONTRIBUTIONS

Kevin J. Lande: Conceptualization (lead); investigation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead).

ACKNOWLEDGMENTS

Thanks to Brandon Ashby, Nick Baker, Lance Balthazar, Jake Beck, Gabe Dupre, Robert Geirhos, E. J. Green, Joan Ongchoco, and two anonymous reviewers for helpful feedback and conversation.

FUNDING INFORMATION

This work was supported by funding from an Insight Development Grant from the Social Sciences and Humanities Research Council of Canada and from the Canada First Research Excellence Fund through York University's "Vision: Science to Applications" program.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Kevin J. Lande https://orcid.org/0000-0002-9543-7787

RELATED WIREs ARTICLES

Computational models of syntactic aquisition
Psychology of knowledge representation
Computational modeling of the mind: What role for mental representation?
Models of visual categorization
Analog mental representation

REFERENCES

Arun, S. P. (2022). Using compositionality to understand parts in whole objects. *European Journal of Neuroscience*, *56*(4), 4378–4392. https://doi.org/10.1111/ejn.15746

Ashby, B. (2020). Rainbow's end: The structure, character, and content of conscious experience. *Mind & Language*, *37*, 395–413. https://doi.org/10.1111/mila.12316

Baker, B., Lansdell, B., & Kording, K. P. (2022). Three aspects of representation in neuroscience. *Trends in Cognitive Sciences*, *26*(11), 942–958. https://doi.org/10.1016/j.tics.2022.08.014

Baker, N., Garrigan, P., & Kellman, P. J. (2020). Constant curvature segments as building blocks of 2D shape representation. *Journal of Experimental Psychology: General*, *150*, 1556–1580. https://doi.org/10.1037/xge0001007

Baker, N., & Kellman, P. J. (2018). Abstract shape representation in human visual perception. *Journal of Experimental Psychology: General*, *147*(9), 1295–1308. https://doi.org/10.1037/xge0000409

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12), 13. https://doi.org/10.1167/9.12.13

Barenholtz, E., & Tarr, M. J. (2006). Reconsidering the role of structure in vision. *Psychology of Learning and Motivation*, *47*, 157–180. https://doi.org/10.1016/s0079-7421(06)47005-5

Bertamini, M., & Wagemans, J. (2013). Processing convexity and concavity along a 2-D contour: Figure-ground, structural shape, and attention. *Psychonomic Bulletin & Review*, *20*(2), 191–207. https://doi.org/10.3758/s13423-012-0347-2

Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: A (re-)emerging program of research for language and vision. *Biolinguistics*, *4*(2–3), 174–200. https://doi.org/10.5964/bioling.8783

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177. https://doi.org/10.1016/0010-0285(82)90007-x

Block, N. (1983). Mental pictures and cognitive science. *The Philosophical Review*, *92*(4), 499–541.

Block, N. (2023). *The border between seeing and thinking*. Oxford University Press.

Blum, H. (1973). Biological shape and visual science (part I). *Journal of Theoretical Biology*, *38*(2), 205–287.

Blum, H., & Nagel, R. N. (1978). Shape descriptions using weighted symmetric axis features. *Pattern Recognition*, *10*(3), 167–180.

Boucart, M., Delord, S., & Giersch, A. (1994). The computation of contour information in complex objects. *Perception*, *23*(4), 399–409. https://doi.org/10.1068/p230399

Burge, T. (2010). Origins of Perception. *Disputatio*, *4*(29), 1–38. https://doi.org/10.2478/disp-2010-0009

Burge, T. (2014). Perception: Where mind begins. *Philosophy*, *89*(3), 385–403. https://doi.org/10.1017/S003181911400014X

Camp, E. (2007). Thinking with maps. *Philosophical Perspectives*, *21*(1), 145–182. https://doi.org/10.1111/j.1520-8583.2007.00124.x

Cao, R. (2022). Putting representations to use. *Synthese*, *200*(2), 151. https://doi.org/10.1007/s11229-022-03522-3

Cavanagh, P. (2021). The language of vision. *Perception*, *50*(3), 195–215. https://doi.org/10.1177/0301006621991491

Chalmers, D. J. (1992). Subsymbolic computation and the Chinese room. In J. Dinsmore (Ed.), *The symbolic and connectionist paradigms: Closing the gap* (pp. 25–48). Lawrence Erlbaum.

Cohen, E. H., & Singh, M. (2007). Geometric determinants of shape segmentation: Tests using segment identification. *Vision Research*, *47*(22), 2825–2840. https://doi.org/10.1016/j.visres.2007.06.021

Cortese, J. M., & Dyre, B. P. (1996). Perceptual similarity of shapes generated from Fourier descriptors. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 133–143. https://doi.org/10.1037//0096-1523.22.1.133

Dakin, S. C., & Baruch, N. J. (2009). Context influences contour integration. *Journal of Vision*, *9*(2), 1–13. https://doi.org/10.1167/9.2.13

Davidson, D. (1967). Truth and meaning. *Synthese*, *17*(3), 304–323.

De Winter, J., & Wagemans, J. (2006). Segmentation of object outlines into parts: A large-scale integrative study. *Cognition*, *99*(3), 275–325.

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., Kording, K. P., Konkle, T., van Gerven, M. A. J., Kriegeskorte, N., & Kietzmann, T. C. (2023). The neuroconnectionist research programme. *Nature Reviews Neuroscience*, *24*(7), 431–450. https://doi.org/10.1038/s41583-023-00705-w

Dretske, F. I. (1981). *Knowledge & the flow of information*. The MIT Press. https://doi.org/10.1086/289062

Elder, J. H. (2018). Shape from contour: Computation and representation. *Annual Review of Vision Science*, *4*, 423–450. https://doi.org/10.1146/annurev-vision-091517-034110

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of gestalt laws for the perceptual organization of contours. *Journal of Vision*, *2*(4), 324–353. https://doi.org/10.1167/2.4.5

Elder, J. H., Oleskiw, T. D., Yakubovich, A., & Peyré, G. (2013). On growth and Formlets: Sparse multi-scale coding of planar shape. *Image and Vision Computing*, *31*(1), 1–13. https://doi.org/10.1016/j.imavis.2012.11.002

Feldman, J. (1997). Regularity-based perceptual grouping. *Computational Intelligence*, *13*(4), 582–623. https://doi.org/10.1111/0824-7935.00052

Feldman, J. (2023). Probabilistic origins of compositional mental representations. *Psychological Review*, *131*, 599–624. https://doi.org/10.1037/rev0000452

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014–18019. https://doi.org/10.1073/pnas.0608811103

Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., & Wilder, J. (2013). An integrated Bayesian approach to shape representation and perceptual organization. In *Shape perception in human and computer vision* (pp. 55–70). Springer. https://doi.org/10.1007/978-1-4471-5195-1_4

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, *33*(2), 173–193. https://doi.org/10.1016/0042-6989(93)90156-q

Fodor, J. (1975). *The language of thought*. Harvard University Press.

Fodor, J., & McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, *35*(2), 183–204. https://doi.org/10.1016/0010-0277(90)90014-B

Garner, W. R. (1974). *The processing of information and structure*. Lawrence Elbaum Associates. https://doi.org/10.4324/9781315802862

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*(6), 711–724. https://doi.org/10.1016/s0042-6989(00)00277-7

Geisler, W. S., & Super, B. J. (2000). Perceptual organization of two-dimensional patterns. *Psychological Review*, *107*(4), 677–708. https://doi.org/10.1037//0033-295x.107.4.677

Gershman, S. J., Tenenbaum, J. B., & J/"akel, F. (2016). Discovering hierarchical motion structure. *Vision Research*, *126*, 232–241.
Gheorghiu, E., & Kingdom, F. A. A. (2007). The spatial feature underlying the shape-frequency and shape-amplitude after-effects. *Vision Research*, *47*(6), 834–844. https://doi.org/10.1016/j.visres.2006.11.023
Graham, P. J. (2014). The function of perception. In *Virtue epistemology naturalized* (pp. 13–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-04672-3_2
Green, E. J. (2020). The perception–cognition border: A case for architectural division. *The Philosophical Review*, *129*(3), 323–393.
Green, E. J. (2023). A pluralist perspective on shape Constancy. *The British Journal for the Philosophy of Science*. Advanced online publication. https://doi.org/10.1086/727427
Green, E. J., & Quilty-Dunn, J. (2017). What is an object file? *British Journal for the Philosophy of Science*, *72*(3), 665–699. https://doi.org/10.1093/bjps/axx055
Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001
Gregory, R. (1970). The grammar of vision. *The Listener*, *83*(2134), 242–244.
Griffiths, T. L., Tenenbaum, J. B., & Kemp, C. (2012). Bayesian inference. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 22–35). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0003
Gross, C. G. (2002). Genealogy of the "grandmother cell". *The Neuroscientist*, *8*(5), 512–518.
Habak, C., Wilkinson, F., Zakher, B., & Wilson, H. R. (2004). Curvature population coding for complex shapes in human vision. *Vision Research*, *44*(24), 2815–2823. https://doi.org/10.1016/j.visres.2004.06.019
Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, *25*(6), 475–492. https://doi.org/10.1016/j.tics.2021.01.006
Hafri, A., Green, E. J., & Firestone, C. (2023). Compositionality in Visual perception. *Behavioral and Brain Sciences*, 46, e277. https://doi.org/10.31234/osf.io/trg7q
Haushofer, J., Baker, C. I., Livingstone, M. S., & Kanwisher, N. (2008). Privileged coding of convex shapes in human object-selective cortex. *Journal of Neurophysiology*, *100*, 753–762. https://doi.org/10.1152/jn.90310.2008
Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Wiley.
Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, *18*(1), 65–96. https://doi.org/10.1016/0010-0277(84)90022-2
Hummel, J. E. (2013). Object recognition. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 32–45). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195376746.013.0003
Hupkes, D., Dankers, V., Mul, M., & Bruni, E. (2020). Compositionality decomposed: How do neural networks generalise? *Journal of Artificial Intelligence Research*, *67*, 757–795. https://doi.org/10.1613/jair.1.11674
Janssen, T. M. V. (2011). Compositionality. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (2nd ed., pp. 495–553). Elsevier.
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211.
Jones, M., & Goldstone, R. L. (2013). The structure of integral dimensions: Contrasting topological and cartesian representations. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 111–132. https://doi.org/10.1037/a0029059
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-o
Keane, B. P. (2018). Contour interpolation: A case study in modularity of mind. *Cognition*, *174*, 1–18. https://doi.org/10.1016/j.cognition.2018.01.008
Keemink, S. W., & van Rossum, M. C. W. (2016). A unified account of tilt illusions, association fields, and contour detection based on elastica. *Vision Research*, *126*, 164–173. https://doi.org/10.1016/j.visres.2015.05.021
Kellman, P. J., & Fuchser, V. (2023). Visual completion and intermediate representations in object formation. In A. Mroczko-Wąsowicz & R. Grush (Eds.), *Sensory individuals: Unimodal and multimodal perspectives* (pp. 55–76). Oxford University Press. https://doi.org/10.1093/oso/9780198866305.003.0004
Kellman, P. J., Garrigan, P., & Erlikhman, G. (2013). Challenges in understanding visual shape perception and representation: Bridging subsymbolic and symbolic coding. In S. J. Dickinson & Z. Pizlo (Eds.), *Shape perception in human and computer vision: An interdisciplinary perspective* (pp. 249–274). Springer-Verlag. https://doi.org/10.1007/978-1-4471-5195-1_18
Kemler Nelson, D. G. (1993). Processing integral dimensions: The whole view. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(5), 1105–1113. https://doi.org/10.1037/0096-1523.19.5.1105
Khuu, S. K., Honson, V., & Kim, J. (2016). The perception of three-dimensional contours and the effect of luminance polarity and color change on their detection. *Journal of Vision*, *16*(3), 1–15. https://doi.org/10.1167/16.3.31
Kimia, B. B. (2003). On the role of medial geometry in human vision. *Journal of Physiology-Paris*, *97*(2–3), 155–190.
Knill, D. C., Kersten, D., & Yuille, A. (1996). Introduction: A Bayesian formulation of visual perception. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 1–21). Cambridge University Press.
Kovács, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure–ground segmentation. *Proceedings of the National Academy of Sciences*, *90*(16), 7495–7497. https://doi.org/10.1073/pnas.90.16.7495
Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253. https://doi.org/10.1017/s0140525x16001837
Lande, K. J. (2021). Mental structures. *Noûs*, *55*(3), 649–677. https://doi.org/10.1111/nous.12324

Lande, K. J. (2023). Contours of vision: Towards a compositional semantics of perception. *The British Journal for the Philosophy of Science*. Advanced online publication. https://doi.org/10.1086/725094

Larson, R., & Segal, G. (1995). *Knowledge of meaning: An introduction to semantic theory*. MIT Press.

Lepori, M. A., Serre, T., & Pavlick, E. (2023). Break it down: Evidence for structural compositionality in neural networks. *Advances in Neural Information Processing Systems*, *36*, 1–38.

Leyton, M. (1988). A process-grammar for shape. *Artificial Intelligence*, *34*(2), 213–247. https://doi.org/10.1016/0004-3702(88)90039-2

Leyton, M. (1992). *Symmetry, causality, mind*. The MIT Press.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, *10*(4), 903–940. https://doi.org/10.1162/089976698300017557

Liu, H., Liu, W., & Latecki, L. J. (2010). *Convex shape decomposition*. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2010.5540225

Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review*, *79*(5), 410–419. https://doi.org/10.1037/h0033129

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. The MIT Press. https://doi.org/10.7551/mitpress/9780262514620.001.0001

McCloskey, M., Valtonen, J., & Cohen Sherman, J. (2006). Representing orientation: A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology*, *23*(5), 680–713. https://doi.org/10.1080/02643290500538356

McDowell, J. (1996). *Mind and world*. Harvard University Press.

McManus, J. N. J., Li, W., & Gilbert, C. D. (2011). Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences*, *108*(24), 9739–9746. https://doi.org/10.1073/pnas.1105855108

Neander, K. (2017). *A Mark of the Mental: In Defense of informational teleosemantics*. The MIT Press.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113–126.

Nguyenkim, J. D., & DeAngelis, G. C. (2003). Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *The Journal of Neuroscience*, *23*(18), 7117–7128. https://doi.org/10.1523/jneurosci.23-18-07117.2003

Pagin, P., & Westerståhl, D. (2010). Compositionality II: Arguments and problems. *Philosophy Compass*, *5*(3), 265–282.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*(4), 441–474. https://doi.org/10.1016/0010-0285(77)90016-0

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Roach & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Lawrence Elbaum Associates.

Partee, B. (2004). Compositionality. In *Compositionality in formal semantics* (pp. 153–181). Blackwell.

Partee, B. (2014). A brief history of the syntax–semantics interface in Western formal linguistics. *Semantics-Syntax Interface*, *1*(1), 1–21.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, *86*(5), 2505–2519.

Peters, B., & Kriegeskorte, N. (2021). Capturing the objects of vision with neural networks. *Nature Human Behaviour*, *5*(9), 1127–1144. https://doi.org/10.1038/s41562-021-01194-6

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*(1), 49–70. https://doi.org/10.1023/a:1026553619983

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65–97. https://doi.org/10.1016/0010-0277(89)90014-0

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2022). The best game in town: The re-emergence of the language of thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences*, *46*, e261. https://doi.org/10.1017/S0140525X22002849

Richards, W., & Hoffman, D. D. (1985). Codon constraints on closed 2D shapes. *Computer Vision, Graphics, and Image Processing*, *31*(3), 265–281. https://doi.org/10.1016/0734-189X(85)90031-3

Rock, I. (1985). *The logic of perception*. The MIT Press.

Sassi, M., Demeyer, M., Machilsen, B., Putzeys, T., & Wagemans, J. (2014). Both predictability and familiarity facilitate contour integration. *Journal of Vision*, *14*(5), 11. https://doi.org/10.1167/14.5.11

Schmidtmann, G., Jennings, B. J., & Kingdom, F. A. A. (2015). Shape recognition: Convexities, concavities and things in between. *Scientific Reports*, *5*, 17142. https://doi.org/10.1038/srep17142

Schwartz, O., & Sanchez Giraldo, L. G. (2017). Behavioral and neural constraints on hierarchical representations. *Journal of Vision*, *17*(3), 13. https://doi.org/10.1167/17.3.13

Schwartz, O., Sejnowski, T. J., & Dayan, P. (2009). Perceptual organization in the tilt illusion. *Journal of Vision*, *9*(4), 19. https://doi.org/10.1167/9.4.19

Shea, N. (2018). *Representation in cognitive science*. Oxford University Press. https://doi.org/10.1093/oso/9780198812883.001.0001

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*(1), 1–23. https://doi.org/10.1017/s0140525x00052432

Spröte, P., Schmidt, F., & Fleming, R. W. (2016). Visual perception of shape altered by inferred causal history. *Scientific Reports*, *6*(36245), 1–11. https://doi.org/10.1038/srep36245

Stevens, K. A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, *46*(3), 183–195. https://doi.org/10.1007/BF00336800

Teller, D. Y. (1984). Linking propositions. *Vision Research*, *24*(10), 1233–1246.

Todd, J. T., & Petrov, A. A. (2022). The many facets of shape. *Journal of Vision*, *22*(1), 1. https://doi.org/10.1167/jov.22.1.1

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

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Vaziri, S., Pasupathy, A., Brincat, S. L., & Connor, C. E. (2009). Structural representation of object shape in the brain. In S. J. Dickinson, A. Leonardis, B. Schiele, & M. J. Tarr (Eds.), *Object categorization: Computer and human vision perspectives* (pp. 182–195). Cambridge University Press. https://doi.org/10.1017/cbo9780511635465.011

Võ, M. L.-H. (2021). The meaning and structure of scenes. *Vision Research*, *181*, 10–20. https://doi.org/10.1016/j.visres.2020.11.003

von Helmholtz, H. (1925). In J. P. C. Southall (Ed.), *Treatise on physiological optics* (Vol. 3). The Optical Society of America.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308.

Zerroug, A., Vaishnav, M., Colin, J., Musslick, S., & Serre, T. (2022). A benchmark for compositional visual reasoning. *Advances in Neural Information Processing Systems*, *35*, 29776–29788.

Zhu, S. C., & Mumford, D. (2006). A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision*, *2*(4), 259–362. https://doi.org/10.1561/0600000018

How to cite this article: Lande, K. J. (2024). Compositionality in perception: A framework. *WIREs Cognitive Science*, e1691. https://doi.org/10.1002/wcs.1691