

UCLA

UCLA Electronic Theses and Dissertations

Title

Parts of Perception

Permalink

<https://escholarship.org/uc/item/4gs9w1bm>

Author

Lande, Kevin Joseph

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Parts of Perception

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Philosophy

by

Kevin Joseph Lande

2018

© Copyright by
Kevin Joseph Lande
2018

ABSTRACT OF THE DISSERTATION

Parts of Perception

by

Kevin Joseph Lande

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2018

Professor C. Tyler Burge, Chair

We perceive the world as organized in certain ways. While a human retina has approximately 100 million photoreceptors, one does not merely see a scene as a collection of 100 million discrete points. Rather, one sees arrangements of complexly shaped, three-dimensional, colored objects—humans, trees, and butterflies, for example. One perceives these things as having parts—limbs, boughs, and wings—each of which has its own specific shape, color, and parts. Not only does one perceive the world as organized in certain ways, one’s very perceptual states are themselves organized from parts. When you perceive a butterfly, the state you are in has as parts your states of seeing the butterfly’s color, texture, and shape. And these states may themselves have parts, though they are not easily identified through introspection. The idea that perceptual states have parts that can combine and recombine in rule-governed ways is central to much of contemporary psychology and over the last century perceptual psychologists have closely investigated the ways in which perceptual states are structured. The aim of this dissertation is to illuminate the relationship between the ways perceptual states are made up from

parts (their *structures*) and the ways that they represent the world as being (their *contents*).

I first examine what it means to say that perceptual states have other perceptual states as parts. I do this by elucidating the explanatory value of attributing part-whole structure to perceptual states. I then defend the principle that perception is *compositional*: the content of a perceptual state depends entirely on the contents of its parts and how those parts are put together. A number of perceptual phenomena, of the sort that interested the early Gestalt psychologists, pose puzzles for the claim that perception is compositional. I show, first of all, how to account for such phenomena within a compositional framework, and, second of all, that compositional accounts of such phenomena are illuminating.

By examining the relationship between the structures of perceptual states and their contents, we can better address central questions in the philosophy of mind. I argue that the structural forms of perceptual states reflect important spatial (and perhaps other) patterns in the perceiver's local environment. This point reveals a fundamental difference between perception and thought. I then address the distinctively perspectival character of perception—the sense in which a slanted coin has an elliptical look even as one correctly sees it as circular. I argue that attempts to explain the perspectival character of perception merely in terms of the properties that one perceives things as having are problematic. Instead, I propose that the perspectival character of perception is rooted in the ways our perceptual states are structured.

This work therefore develops a framework for understanding perception by examining the nature of the parts of perception and the ways in which those parts combine to give rise to a rich perception of the world.

The dissertation of Kevin Joseph Lande is approved.

Gabriel Jae Greenberg

Phil Kellman

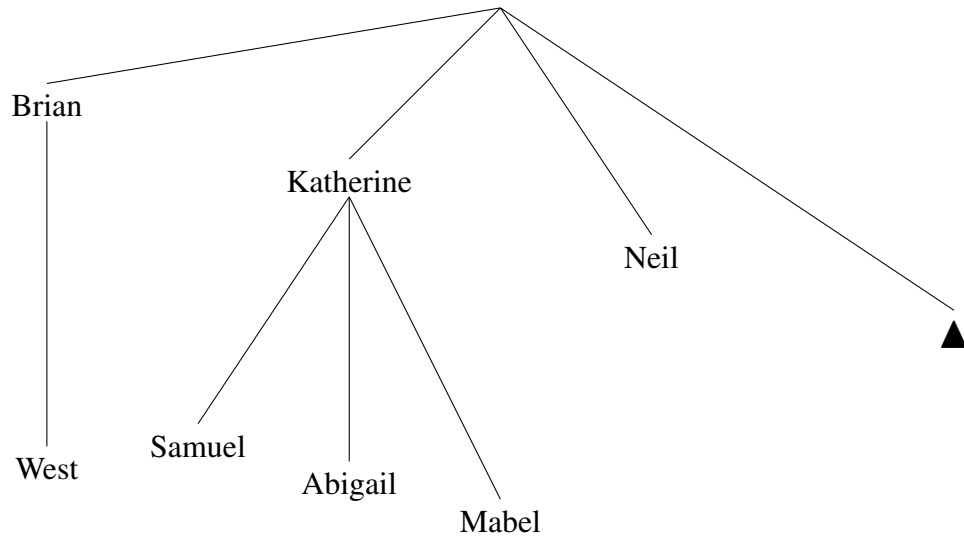
C. Tyler Burge, Committee Chair

University of California, Los Angeles

2018

For:

Gary & Nancy



One wants to say: ‘If our psychology is, in general, right then the nature of the mind must be, roughly, this . . .’ and then fill in the blank. . . . [T]he experimentalist can work the other way around: ‘If the nature of the mind is roughly . . ., then our psychology ought henceforth to look like this: . . .’, where *this* blank is filled by new first-order theories. We ascend, in science, by tugging one another’s bootstraps.

Jerry Fodor

THE LANGUAGE OF THOUGHT

‘Like’ and ‘like’ and ‘like’—but what is the thing that lies beneath the semblance of the thing? . . . There is a square; there is an oblong. The players take the square and place it upon the oblong. They place it very accurately; they make a perfect dwelling-place. Very little is left outside. The structure is now visible; what is inchoate is here stated; we are not so various or so mean; we have made oblongs and stood them upon squares. This is our triumph; this is our consolation.

Virginia Woolf

THE WAVES

TABLE OF CONTENTS

Acknowledgments	xi
Introduction	1
1 Parts of Perception	8
1.1 Representations and Their Parts	12
1.2 The Dynamics of Psychological States	19
1.3 The Demographics of Psychological States	28
1.3.1 Range	32
1.3.2 Distribution	44
1.3.3 Summary	55
1.4 The Generality of Structural Theories	59
1.5 An Independent Channel of Evidence	63
1.6 Psychosyntax and Psychosemantics	68
2 Composing Percepts	72
2.1 Compositionality	75
2.2 The Causal Order and the Semantic Order	82
2.3 Emergent Content	93
2.4 Context Effects	96

2.4.1	Context Effects in Language	98
2.4.2	Lightness and Color	100
2.4.3	Three-Dimensional Edges	107
2.4.4	Orientation and Shape	116
2.5	Compositionality as a Research Program	125
3	The Semantic Significance of Perceptual Structures	129
3.1	What Structure Alone Contributes	132
3.2	Visual Arrays	142
3.3	Feature Binding	151
3.4	Configural Binding	162
3.5	Visual Arrays, Revisited	173
3.6	Form and Function	177
3.7	Forms of Mind	181
4	The Perspectival Character of Perception	186
4.1	Slanted Coin Arguments	189
4.1.1	The Argument from Variance	191
4.1.2	The Argument from Similarity	194
4.2	The Pluralist Approach	196
4.3	The Perspectival Properties Approach	200
4.4	The Argument from Functional Redundancy	205
4.5	The Structures of Perspectival Representations	212
4.6	Conclusion	222
	Bibliography	223

LIST OF FIGURES

1.1	An interpretation of (Palmer, 1977)’s “part verification” experiment.	21
1.2	Stimuli from (Shepard and Metzler, 1971)’s mental rotation experiment.	24
1.3	A “nonsense” shape that one has likely never seen before and that one likely cannot classify under any familiar categories. (From Biederman, 1987, p. 116).	33
1.4	Examples of stimuli used in (Schacter et al., 1991)’s study of priming for possible and impossible objects.	41
1.5	Stimuli from (Biederman and Cooper, 1991)’s study on the role of three-dimensional parts in perceiving more complex three-dimensional shapes.	46
1.6	The very same figure may be perceived as a square or a diamond.	52
1.7	An illustration of a reference frame-based shape representation. (From Hinton and Parsons, 1988, p. 2).	54
2.1	A “nonsense” shape.	78
2.2	Examples of interdependent processing in the perception of (a) collinear line segments (Polat and Sagi, 1993), (b) closed contours (Kovács and Julesz, 1993), and (c) three-dimensional objects (Williams and Weisstein, 1978).	85
2.3	“Hierarchical stimuli.” (From Navon, 1977).	89
2.4	Illustrations of the emergent perception of (a) axes of symmetry, and (b) closure.	94
2.5	Context effects in lightness perception: (a) The simultaneous contrast effect, and (b) the corrugated plaid illusion (Adelson, 2000).	101
2.6	Context effects in the perception of three-dimensional corners.	108

2.7	Necker cube.	112
2.8	Context effects in the perception of orientation: We perceive a triangle as pointing in different directions depending on how it is embedded in a surrounding configuration of triangles. (From Palmer, 1980).	116
2.9	Context effects in shape perception: We perceive the very same figure as either a square or a diamond, depending on how it is situated within an array of other figures.	121
2.10	An illustration of a reference frame-based shape representation. (From Hinton and Parsons, 1988, p. 2).	124
3.1	A conventional map of part of Los Angeles.	139
3.2	An Euler diagram, depicting <i>A</i> as a subset of <i>B</i>	141
3.3	An illustration of a simple scene and an array-like representation of that scene.	143
3.4	An example of feature binding. Under normal circumstances, one will perceive a red circle besides a blue square. Under certain circumstances, one may inaccurately perceive a red square beside a blue circle.	154
3.5	Illustrations of perceptual grouping principles. (From Palmer, 1999, p. 258).	164
3.6	Examples of stimuli from (Palmer, 1977)'s study of perceptual organization.	166
3.7	An interpretation of (Palmer, 1977)'s "part verification" experiment.	168
3.8	An illustration of how feature binding and configural binding operate within a basic array-like structure.	175
4.1	Visual angle.	198
4.2	An illustration of projective shape.	203
4.3	Rough illustrations of array representations as of (a) a head-on circular surface, (b) a slanted circular surface, and (c) a head-on elliptical surface.	215

ACKNOWLEDGMENTS

I would like to begin by thanking my committee. I am especially grateful to my advisor, Tyler Burge, for his guidance and encouragement. He has helped me to see the value in my work even when I, in moments of self-doubt, had lost sight of it. His emphasis on doing philosophy with total seriousness and deep carefulness has been a model for me. I can hardly identify all the insights and skills I have gained from my conversations and seminars with him, and from thinking through his works.

In Gabe Greenberg, I have always found someone who is just as interested in and puzzled by the issues that interest and puzzle me. On many occasions, he patiently supported me as I tried to make my way through those issues. My discussions with Gabe have always filled me with a sense of clarity and energy for doing philosophy. Sam Cumming has regularly motivated me not to rest content in approaching a problem with one set of techniques and from the perspective of one field of inquiry, but to approach a problem using any technique that might be illuminating and with help from any field that might be relevant. Finally, in addition to benefiting me with his expertise, every interaction with Phil Kellman reminds me of what a pleasure it is to think with others about philosophy and psychology.

I benefited greatly from conversations with Josh Armstrong, Ned Block, Ian Boon, Denis Buehler, Gabe Dupre, Katie Elliott, E.J. Green, Gabby Johnson, Bill Kowalsky, Jake Quilty-

Dunn, Susanna Schellenberg, Sheldon Smith, and Tamar Weber. Special thanks to the members of the UCLA Mind and Language Workshop for bearing through my many presentations and re-presentations of the material in this dissertation.

The UCLA Philosophy Department has transformed how I do philosophy, both on my own and in conversation with others. I have been reinforced with the need (not necessarily satisfied here) to always ensure that the concepts, distinctions, arguments, and techniques employed in philosophy actually connect to the heart of the matter. Thanks especially, in this respect, to Seana Shiffrin, Tyler Burge, David Kaplan, Barbara Herman, and Pamela Hieronymi.

While writing the dissertation that follows, I often turned to the same texts for inspiration. Jerry Fodor's work, and in particular *The Language of Thought* (1975), has served as a standard, for the way he integrated empirical psychology with formal concepts from the study of language, logic, and computation, so as to illuminate deep philosophical problems. And, throughout, I have continued to learn new and rich lessons from Tyler Burge's *Origins of Objectivity* (2010). I have continually returned to the early exemplars of rich research programs. These include Stephen Palmer's theoretical attempt to regiment the notion of cognitive representation (Palmer, 1978), as well as his empirical work on perceptual organization (Palmer, 1977); David Marr's and Zenon Pylyshyn's defenses of computational cognitive science (Marr, 1982; Pylyshyn, 1984); the linguistic program of generative grammar, especially as advertised in Noam Chomsky's early work (Chomsky, 1957, 1965); model-theoretic formal semantics of natural languages, emerging from the work of Richard Montague and Barbara Partee (Montague, 1970, 1973; Partee, 1975, 2004); work on machine vision and pattern recognition that attempted to marry the conceptual tools from computational theory, gener-

ative syntax, and formal semantics with the domain of visual perception (Sutherland, 1968; Clowes, 1971; Fu, 1974; Reiter and Mackworth, 1989); and the literature on iconic representations (especially Camp, 2007; Rescorla, 2009; Shimojima, 2001; Greenberg, 2011), which has offered fruitful models for applying the tools of generative syntax and formal semantics to understanding non-linguistic representations. It has not escaped me that there is a dearth of women and people of color on this list. In fact, a glance at the bibliography will show a disappointing underrepresentation of these groups. This is due to a combination of reprehensible facts about many of the fields from which I have drawn, as well as my failing (reprehensibly) to correct for the inadequacies of my methods for discovering literature. I promise to do better.

I would not be where I am without Gordon Brittan's kindness and mentorship during my time as an undergraduate at Montana State University, and Jerry Samet's guidance during my time as an M.A. student at Brandeis University. I am grateful to Greg Antill, Joshua Blanchard, Ashley Feinsinger, David Friedell, Laura Gillespie, Jonathan Gingerich, Michael Hansen, Antti Hiltunen, Melissa Hughs, Skef Iterum, Andrew Jewell, Lauren Leydon-Hardy, Derek Leonard, Thi Nguyen, John Nosan, and Eric Tracy, in addition to everyone else mentioned above, for their friendship and support—which is all that really matters anyway. I am particularly grateful to Mariko Green, who was a source of companionship and wonderfulness these last two years. Finally, at my core I am what I am due to my parents, Gary and Nancy, and my siblings, Brian, Katherine, and Neil, all of whom raised and taught me, and who will always be my compass.

Please keep in mind that any inaccuracies in this work are entirely the fault of the facts being what they are.

VITA

- 2015 C.Phil in Philosophy, University of California, Los Angeles.
- 2012 M.A. in Philosophy, University of California, Los Angeles.
- 2011 M.A. in Philosophy, Brandeis University.
- 2009 B.A. in Philosophy (Highest Honors), B.A. in Political Science (Highest Honors), and University Honors Baccalaureate (Highest Distinction), Montana State University.

PUBLICATIONS AND PRESENTATIONS

“The Perspectival Character of Perception.” (Forthcoming). In *The Journal of Philosophy*.

“Consciousness and the Format of Perception,” The Quale 3rd Interdisciplinary Academic Symposium, at UCLA; November 2017.

“Percepts, Pictures, and Their Parts,” Representation, Meaning, and Content Workshop, at the

Center for Logic, Language, and Cognition, University of Turin; July 2017.

“Do We Perceive Perspectival Properties?” 2nd IIFs-UNAM Philosophy Graduate Conference, at Universidad Nacional Autónoma de México; March 2016.

“Is Perception Compositional?” Semantics and Philosophy in Europe (Eighth Colloquium), at University of Cambridge; September 2015.

“The View from the Great Wave,” UCLA-LACMA Symposium: Art Historical Theories and Methodologies, at the Los Angeles County Museum of Art; March 2015.

“Computation, Representation, and Empirical Consequences” (poster), Society for Philosophy and Psychology, at University of British Columbia; June 2014.

“Realizing Functions: Functionalism and Indeterminacy,” Interdisciplinary Graduate Conference on Consciousness, at Boston University; May 2011.

“Realizing Functions: Functionalism and Indeterminacy,” 4th Annual Graduate Philosophy Conference, at Western Michigan University; December 2010.

“Quine and the Revisability of Logic: The Revisability Paradox and Logical Implication,” Conference on Foundations of Logical Consequence, at Arché Philosophical Research Centre for Logic, Language, Metaphysics and Epistemology, University of St. Andrews; June 2010.

INTRODUCTION

We perceive the world as organized in certain ways. The scene in front of you consists of arrangements of objects, each having shapes, sizes, colors, textures, and so on. A large part of the philosophy of perception is dedicated to better understanding how perception delivers the complex world that one represents. At the same time, there is good reason to think that perceptual states are themselves complex. You do not just perceive a table as having parts. When you perceive the table, the state you are in itself has parts—states of perceiving the sizes, shapes, and colors of the tabletop and of the legs. These perceptual states themselves have parts. The idea that perceptual states have parts that can combine and recombine in rule-governed ways pervades much of contemporary psychology. Over the last century, perceptual psychologists have closely investigated the ways in which our perceptual states are structured.

The aim of this work is to illuminate the relationship between these two aspects of perception: between the *contents* of perception—the ways that we perceptually represent the world as being—and the *structures* of our perceptual states themselves. I look at some challenges to understanding this relationship, and then show that by spelling out the relationship we can better address central problems in the philosophy of mind. While I focus almost exclusively on human vision, since that is where the psychological literature is the largest, I expect the central claims here to generalize, with suitable modifications, across modalities and animals.

The first half of this work is dedicated to addressing challenges that arise when attempting to articulate how the structures of our perceptual states relate to the contents of those states. One challenge is to understand what it even means for a perceptual state to be structured out of other perceptual states. In Chapter 1 (“Parts of Perception”), I take a non-reductive approach to this problem. Instead of trying to say *what it is* for a perceptual state to be structured, I examine what one can *explain* by attributing structure to a perceptual state. In doing so, I argue that psychologists and philosophers, in explicating the notion that mental states can have part-whole structure, commonly overlook central empirical and explanatory grounds for attributing such structure.

In Chapter 2 (“Composing Percepts”), I defend the principle that perception is *semantically compositional*. According to this principle, the content of a perceptual state is entirely a function of the contents of its constituent parts, together with the way in which those parts are combined (their “mode of combination”). The contribution of this chapter consists not so much in a positive argument for this general principle, but rather in the attempts to show how the principle applies to a number of ubiquitous perceptual phenomena. A great deal of psychological research suggests that our representations of objects and surfaces have as parts representations of smaller fragments of the world. Yet the early Gestalt psychologists already showed that it would be no simple task to explain how our rich perception of discrete objects and groups, with holistic properties such as shape, orientation, and symmetry, could emerge from more these more basic representations of smaller bits of the world. As Max Wertheimer (1938a) suggested, we cannot analyze perceptual representations as mere “and-summations” of causally and structurally independent percepts.

I explain how to make sense of three facts that pose puzzles for developing a compositional semantics for perception. First, one's representations of the larger, more complex parts of the world seem to have some causal—not to mention phenomenological—priority over one's representations of the smaller bits. Second, representations of whole objects and surfaces seem to carry rich kinds of content that are not carried by component representations. Plausibly, one's representation of a square has as parts representations of the sides of that square. While one does not perceive those sides as having properties such as *closure* and *area*, one does attribute such properties to the square as a whole. Finally, *context effects*, in which the way you perceive one element of a scene depends on your perception of other elements in a scene, are ubiquitous in perception.

I show how these facts can be handled by compositional accounts of perceptual phenomena, arguing that there is a feasible and illuminating research program in developing and testing such accounts. The principle of compositionality offers a conceptual framework for identifying puzzles and for developing empirically testable solutions to those puzzles. The principle forces one to regiment one's understanding of *content*, *structure*, and *process* in perception. Very often in philosophy and psychology, these concepts are not suitably distinguished. When those concepts are not adequately distinguished, it is all the more difficult to understand how they relate to each other in particular cases.

The work in these first two chapters is as much philosophy of science as it is philosophy of mind. I hope that these chapters may be of service to the working psychologist. I attempt to apply insights from the empirical study of syntax and formal semantics in linguistics toward understanding perception. While perception differs quite radically from natural language, I

believe we can learn a lot about how to study perception by looking at how linguists have studied natural languages. Psychological and philosophical attempts to theorize about the concepts of perceptual structure and content have drawn a great deal from mathematical logic and the formal theory of computation. What is often lost in these discussions is an understanding of how one actually goes about empirically discovering the structures and contents of perceptual states and how one puts those discoveries to explanatory use. I hope, in these first two chapters, to better illuminate how the tools provided by the formal study of representation can be applied empirically in perceptual psychology. With this aim in mind, I have chosen to prioritize discussions of empirical methodology and conceptual explication over giving systematic formalizations of the structure and semantics of perceptual states. Moreover, in my discussion of the empirical literature I prioritize what I take to be exemplars of certain explanatory frameworks or methodologies over more refined, contemporary instances. I have nevertheless made an effort to indicate the current state of the art where it is helpful to understanding certain explanatory or methodological strategies and their ongoing utility, or where I make specific factual claims.

In the latter half of the dissertation, I show how we can get greater purchase on central issues in the philosophy of mind by studying how the structures of our perceptual states relate to the contents of those states. A central project in the philosophy of mind is to better elucidate the deep differences between the various mental capacities we find in nature. How, for instance, does perception differ from high-level human thought? We can make progress in our attempt to articulate the natural joints of the mind by examining the semantic commitments of different types of mental representational structure. In Chapter 3 (“The Semantic Signifi-

cance of Perceptual Structure”), I argue that the very structural forms of visual states commit those states to being about specific kinds of spatial arrangements. I argue that different types of visual structure carry different kinds of spatial commitments. These types of structure are all rooted, I argue, in underlying array-like structures. Roughly, the thought is that for a core class of visual representations, the organization of the visual array—the array of lines of sight from one’s viewpoint to things in the world—is embodied in the structural forms of those representations. Many other types of visual states are structured from these basic array-like representations.

The main burden of Chapter 3 is to argue that different types of spatial arrangements, including but not limited to the arrangement of the visual array, are not merely encoded in perception, but are entailed by the very structural forms of perceptual states. I suggest that this claim is an instance of a more general regularity: throughout different perceptual modalities in different species, the structural forms of perceptual states reflect ecologically significant types of arrangements—arrangements the tracking of which is important to the perceiver’s survival or to her veridically perceiving other important features of the environment.

It is plausible that structural forms of human thought lack the specialized kinds of semantic commitments that are carried by perceptual structures. While a core function of any given perceptual capacity is to represent important aspects of one’s surrounding physical environment, a function of human thought is, in part, to flexibly integrate content about disparate kinds of topics. The semantic commitments of perceptual structure facilitate the specialized functions of perceptual capacities. By contrast, the generality of thought means that the structural forms involved must have far more abstract semantic commitments. Perception and

thought may well lie at distant ends of a spectrum—with the structural forms of perception being highly specialized, while the structural forms of thought are highly general in their applicability.

In Chapter 4 (“The Perspectival Character of Perception”), I look at attempts to reconcile the objectivity of perception with its perspectival character. One can perceive things, in many respects, as they really are. Nonetheless, one’s perception of the world is perspectival. For example, while you can correctly see a coin as circular from most angles, the coin looks different when slanted than when head-on, and in fact there is some respect in which the slanted coin looks similar to a head-on ellipse. If perception is objective, why is it also perspectival? Many hold that perception is perspectival because we perceive certain properties that correspond to the “looks” of things. This sort of approach attempts to explain the perspectival character of perception solely in terms of the denotational contents of perceptual states. I argue that this approach is misguided. I consider the two standard versions of the approach. What I call the *pluralist approach* fails to give a unified account of the perspectival character of perception, while what I call the *perspectival properties approach* violates central commitments of contemporary psychology.

I propose instead that perception is perspectival because of the way perceptual states are structured from their parts. The visual representation of a slanted coin may not attribute any relevant properties in common with the visual representation of a head-on ellipse. Rather, I suggest that the structural forms of those representations are similar. If the visual states are array-like, for instance, then the two representations will be organized into the same patterns in the representational array. More generally, perceptual states exhibit structural interdepen-

dencies that reflect the ways that properties in the world get conflated by our sensory stimuli. For example, the perception of shape and the perception of orientation are structurally interdependent in array-like representations. The specific ways in which these perceptual representations are interdependent give rise to perspectival phenomena—to structural similarities, for instance, between the visual representation of a slanted coin and the visual representation of a head-on ellipse.

This dissertation therefore develops a framework for understanding perception by examining the nature of the parts of perception and the ways in which those parts combine to give rise to a rich perception of the world. Each of these chapters progressively enriches and applies this overall framework. At the same time, each chapter is concerned with a largely independent set of claims. One should in principle be able to read each chapter on its own.

CHAPTER 1

Parts of Perception

Kevin Lande: Dear Professor Chomsky,

What would you offer as the most general definition of being a syntactic constituent of a representation. . . ?

Noam Chomsky: Too general for a useful response[.]

PERSONAL CORRESPONDENCE

We perceive the world as organized in certain ways. Consider vision. A human retina has approximately 100 million photoreceptors. But we do not perceive scenes as simply containing 100 million discrete points. We perceive scenes as containing complex three-dimensional shapes—the shapes of humans, trees, and butterflies, for example—that are arranged in space around us. We perceive these three-dimensional shapes as having simpler shapes as parts—the shapes of limbs, boughs, and wings, for example. We perceive these simpler shapes as having textured, colored surfaces that are bounded by contours of certain shapes. So, we perceive structure in the world. A great deal of psychological research over the last half century or so has suggested that the states (or events, or acts) of perception themselves are structured. My state of representing the butterfly's shape is made out of other states—including, perhaps, the states of representing the shapes of the butterfly's wings.

I intend to elucidate what it means to say that perceptual states are structured from or “made out of” other perceptual states. One can think of the perceptual system as endowing us with a number of different perceptual capacities or abilities (see Geach, 1957; Evans, 1982; Burge, 2009)—for example the ability to represent particular shapes, sizes, colors, and so on. We might paraphrase the claim that perceptual states can be structured by saying that perceptual abilities can be structured complexes of other abilities. My abilities to perceive the shapes of the butterfly’s wings are, perhaps, parts of my ability to perceive the shape of the whole butterfly. One might say that a perceptual *state* is complex in case one’s being in that state is the result of exercising a complex perceptual ability. The question of the paper, then, can be paraphrased as the question of what it means to say that perceptual abilities are made up of other perceptual abilities.

I will take a non-reductive approach to addressing this question. Rather than providing an account of what it is for a perceptual state to have parts, I will characterize what one can explain by attributing structure to perceptual states and, by extension, to other types of psychological states. My aim here is not to persuade skeptics that psychological states in general, or perceptual states in particular, can be structured. Nevertheless, I expect that elucidating the explanatory value of attributing part-whole structure to perceptual states, and showing how such attributions are used throughout perceptual psychology, can serve as part of a case that perceptual states really can have such structure.

Perceptual psychology is replete with attempts to characterize the constituent ingredients of perceptual states. Some have proposed, for example, that our perception as of the shapes of objects are made out of our perceptions as of of the three-dimensional shapes of the parts of

those objects (Marr and Nishihara, 1978; Biederman, 1987). Others have disagreed, arguing, for example, that our visual perceptions as of scenes and the shapes of objects are made out of perceptions as of the *two-dimensional* shapes of visible surfaces and not the three-dimensional shapes of the volumes of objects and parts (Nakayama et al., 1995; Leek et al., 2005, 2009). Some have argued that when we visually perceive the shape of an object's outline, our perceptual state is made up of states of seeing individual segments of that outline (Attneave, 1954; Sutherland, 1968; Leeuwenberg, 1971; Richards et al., 1986; De Winter and Wagemans, 2006; Garrigan and Kellman, 2011). Some theories of perceptual organization suggest that the structure that we perceive a scene to have is reflected in the part-whole structure of our perceptual states (Palmer, 1977).

The notion of part-whole structure that is used in psychology is associated with a notion of structural possibility—that is, the notion that there may be constraints on how parts can combine into a whole. Hence, in addition to specifying what kinds of parts a type of perceptual state has, structural theories in psychology may also specify the constraints on how those parts can and cannot combine to form the more complex type of state. Some theories suggest, for example, that in order for visual perceptual representations of shapes to combine into complex representations of more articulated shapes, the former representations must be as of shapes that intersect with each other, forming points of sharp concavity (Hoffman and Richards, 1984; Singh and Hoffman, 2001). There is a great deal of research suggesting that visual perceptual representations of segments of contours can combine into complex representations of more elongated contours—and ultimately, into representations of the bounding outlines of objects—only if, among other things, the former representations are as of contour

segments that are sufficiently collinear and close to each other (Palmer, 1977; Feldman, 1997; Geisler and Super, 2000; Elder, 2015). Studies of perceptual grouping typically look at what characteristics representations as of individual elements must have in order to be integrated into a complex representation as of a group of elements.

I will examine what these psychologists are trying to explain when they attribute part-whole structure to perceptual states. My aim is not just to describe an existing concept of psychology but also to clarify that concept so that it can be used more precisely in future theoretical and empirical work. I begin, in Section 1.1, by drawing a preliminary distinction between the kind of part-whole structure in which we are interested, which I call “representational structure” or “syntactic structure,” and other kinds of structure. In Section 1.2, I discuss the role that representational structure plays in explaining the *dynamics* of our psychological states—that is, the causal and temporal relations between those states. Since structural theories can generalize across different stories about how psychological states are causally connected, those theories must have other more general *explananda* as well. In Section 1.3, I describe a number of cases in which psychologists theorize about representational structure in order to explain what I call the *demographics* of our psychological states, including their “range” and “distribution.” The *range* of psychological states concerns the law-like constraints on what psychological states can or cannot, in principle, occur. The *distribution* of psychological states concerns the law-like patterns in how psychological states can, cannot, and must co-occur. In Section 1.4, I argue that in explaining the range and distribution of psychological states, structural theories can abstract from, or generalize across, particular theories of psychological processing. In Section 1.5, I appeal to the generality of structural theories to challenge

the widespread view that structural theories can only be evaluated in the context of processing theories. I argue that structure theories can be confirmed or disconfirmed independently of any accompanying processing theories. Structural theories in psychology make their own explanatory and empirical contributions. I conclude, in Section 1.6, with a discussion of the relationship between our theories of the structure that we perceive the world to have and our theories of the structure of our perceptual states themselves.

1.1 Representations and Their Parts

I will use the term “representational structure” to capture the relevant kind of structure that psychological states might have by virtue of being made out of other psychological states. I think one can equally well use the terms “syntactic structure” or “constituent structure.” The concept of representational structure is one in a set of interrelated concepts: *constituents* (or parts), *mode of combination*, *structural constraints*, *structural categories*, and *form*. One identifies the structure of a representation by specifying (a) what its *constituent parts* are (including what their structural categories are), and (b) in what way those parts are combined (their mode of combination). The structural category of a representation determines how it can and cannot combine with other representations. Conversely, for each way of combining some representations—for each mode of combination—there will be certain constraints on what representations (of what structural categories) can and cannot be combined in that way. The form of a representation abstracts away from the specific parts of the representation. A representation’s form is determined by how it is combined from basic parts of certain structural

categories.¹ Representational structure is a formal property, in that it specifies the form of a representation, or how a representation can be put together with other representations and how the representation itself is put together.

It is not enough for our interests, however, to describe how these concepts of structure, constituency, combination, and so on, are interrelated. I intend to elucidate how these inter-related concepts purport to describe a psychological reality. The first step will be to look at the kinds of things to which psychologists apply these concepts. I will then consider some simple attempts to reduce the concept of representational structure to more familiar kinds of structure. In the end, I will propose that we take a more non-reductive approach: instead of saying what representational structure *is*, I will look at what the concept of representational structure *explains*.

Psychologists paradigmatically attribute representational structure to psychological representations. So let's begin by characterizing the notion of a *representation*. Roughly, a *representation* is an object, state, or event that has content (in the sense that one speaks of the content of a speech or report) about how the world, or some aspect of the world, is. Representations can single out particulars and attribute properties to those particulars by virtue of their content.²

We ordinarily describe an experience by describing aspects of its content—by describing

¹My usage of the term “structure” throughout this work is ambiguous. At times I will be referring to the more abstract structural form of a representation—the way the representation is combined from its basic parts, abstracting away from what those specific parts are. At other times, I will be referring to the structural form of a representation *together with its specific parts*. Context should be sufficient to distinguish which of these I mean.

²A “denotational semantics” says what particulars, properties, or states of affairs different representations function to represent.

the way the world is according to that experience. For example, to describe my visual experience at the time of writing I might say that I see the top of my desk as a brown, rectangular surface. The content of a perceptual state determines the circumstances in which that state is accurate or inaccurate. If the the top of the desk is not brown and rectangular, then my experience is inaccurate. I might mention other features of the experience besides its content. For example, I might note how blurry my experience is, since I am not wearing glasses. But I would be severely underdescribing the experience if I did not say something about its content—about how it represents the world.

This practice of characterizing perceptual experiences in terms of their content, or how they represent the world to be, has been made empirically rigorous in the context of perceptual psychology. At the core of perceptual psychology are hypotheses about how certain patterns of behavior depend on our capacities to represent certain features of the world. For example, my picking up a cup depends, among other things, on my capacity to represent the shape and size of that cup. My tossing a crumpled up piece of paper into a trashcan depends, among other things, on my capacity to accurately represent the distance between me and the trash can. Our representational capacities are grounded in complex processes that rely on information about our own physiological states (for example, the activity of photoreceptors on the retina) in order to produce states that represent features of the distal world. A good deal of what is known as *psychophysics* is dedicated to experimentally determining what properties an organism can represent in perception. And a good deal of computational psychology is dedicated to determining how we are able to represent those properties.

Many psychological theories hold that psychological representations can be made of other

constituent representations. By way of analogy, the sentence “Alma laughs” is a linguistic representation made up of or structured from the constituent linguistic representations “Alma” and “laughs.” The *representational structure* of a representation specifies the constituent representations that make up that representation and the way in which those constituents are related so as to make up that representation. Jerry Fodor (1975) argued extensively that the most successful theories in cognitive psychology are committed, on the face of it, to characterizing cognitive states as structured representations. For example, one might characterize my belief that Alma laughs as being made out of a concept of *Alma* and a concept of *laughing*. While Fodor focused on cognition, perceptual psychology is full of theories concerning how perceptual representations are made up of other perceptual representations.

Representations—the objects, states, or events that have content about the world—can have different kinds of parts. If a representation is a material object, then that object may have material parts. If the representation is a temporally extended event—a sequence of sounds for example—then it can have temporal parts. Such material and temporal part-whole relations can be found in many kinds of things, both representational and non-representational. But representational structure is a kind of part-whole structure that is paradigmatically found in objects, states, or events that function to represent.³ What is this particular kind of part-whole structure?

³As I use the term, “representational structure” refers to part-whole structure of the representation itself—the object, state, or event that has content. I reserve the term “represented structure” to refer to whatever structure the representation represents the world as having.

It may be helpful to think of a representation’s content as having a part-whole structure that mirrors the structure of the representation, by analogy with the idea that sentences have *structured meanings* that reflect the syntax of those sentences (see Lewis, 1970; Cresswell, 1985; Soames, 1987). We may reserve the term “semantic structure” to refer to the structure of the content by which the representation functions to represent what it does.

One can find many metaphorical descriptions of representational structure as the “shape” of a representation (Fodor, 1983) or “the nature of the marks used in the representation (such as ink, magnetic fluxes, or sound waves)” (Kosslyn, 1981, p. 46). Some describe representations as “physical patterns,” while complex representations are “composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another)” (Newell and Simon, 1976, p. 116). These descriptions easily risk suggesting that representational structure must be defined in terms of other natural kinds of part-whole structure. On such a view, what it is to be a constituent of a representation or to be a representational structure is at least partially to be a certain independently specified, non-representational natural kind of part or part-whole relation. For instance, one might hold that representational constituents must be identifiable with natural kinds of physical or material constituents.

One problem with this view is that token instances of the same kind of representational structure can differ widely in their non-representational structures. The same type of perceptual state with the same type of representational structure could occur in a human and a Martian whose neural events have very different neurophysiological structures. A spoken utterance and a written inscription of the same type of sentence structure may have very little in common by way of material, spatial, and temporal part-whole structure. It is unlikely that there is any natural kind of non-representational part-whole structure that is common to every instance of a type of representational structure (cf. Putnam, 1967; Fodor, 1974).

A distinct problem with the view is that where representational structure *supervenies* on some non-representational structure, that non-representational structure is unlikely to be a *natural kind* of structure independently of the fact that it underlies a kind of representational

structure. The physical patterns that correspond to a particular utterance's having a certain syntactic structure may not be natural kinds of patterns. Or if these patterns are natural kinds, this depends on the fact that they ground utterances with that kind of syntax. Even if the representational structure of a psychological state supervenes on the structure of an underlying brain state, there is little reason to expect that the relevant structure of the brain state would correspond to any natural kind of structure, or at least any kind independently specified in neuroscience.

A related *epistemic* point is that types of representational structure in psychology are not typically posited by reference to types of non-representational structure. Rarely do psychologists define kinds of representational structure in terms of kinds of neural structure. One does not typically look at the physical patterns in the brain, for example, and extract the representational structure of the psychological state. Psychological theories of representational structure are typically built around behavioral evidence from psychophysics. If anything, psychologists define natural kinds of non-representational, neural structure on the basis of their correlations to types of representational structure that have been posited independently on the basis of behavioral evidence.

It is also common to gloss structural properties as being the "causal roles" of psychological states (see, for example, Field, 1978; Stich, 1983). On a simple form of causal role functionalism about representational structure, all there is to having a given representational structure is entering into a certain system of causal relations. As with functionalist theories in other domains, we run into difficulties as soon as we try to say what set of potential causes and effects are necessary and sufficient for having a certain kind of representational part-whole

structure (cf. Block, 1980). The problem is especially difficult because psychological hypotheses about representational structure are often noncommittal and open-ended about the causes and effects of such structure. The claim that our perceptual representations of humans, trees, and butterflies are structured from perceptual representations of (among other things) limbs, boughs, and wings is in principle compatible with a wide variety of stories about how these representations are causally related to each other and to other states. So we must not identify kinds of representational structure with very specific causal roles. On the other hand, we must not identify structural kinds with causal roles that are too general. Defining a structural kind as *whatever property causally explains such-and-such phenomena*, for example, risks rendering appeals to that structural kind as explanatorily vacuous. So it is not clear that this simple form of functionalism gives us much purchase on the nature of representational structure.⁴

I propose to treat representational structure as a *sui generis* kind of part-whole structure that is not type-identical with other kinds of part-whole structure or with causal roles. Representational structural properties certainly depend for their instantiation on the instantiation of physical properties or relations. But I believe we are far from having any adequate account of that dependence relation.

How, then, can we understand what representational structure is if not by relating it to some antecedently understood kind of structure? I propose that rather than trying to relate

⁴I do not mean to reject the claim that a state's representational structure helps to determine, and is in part determined by, that state's causal dispositions. However, I am skeptical that types of representational structure can be *identified* with systems of causal dispositions. I also do not mean to rule out the possibility one could attribute to psychological states a kind of abstract structure that is type-identified by its specific causal role. But that notion of structure is not the notion being used in the psychological theories discussed here. Such a notion of abstract causal structure cannot prescind from processing theories in the way that psychological theories of representational structure commonly do.

representational structure to other kinds of structure, we look at what sorts of things representational structure can explain. I should note that for the purposes of this discussion I am not concerned with whether specific hypotheses about representational structure are true or not. Instead, I am interested in the grounds according to which psychologists judge hypotheses of representational structure to be explanatory (or not) and justified (or not).

1.2 The Dynamics of Psychological States

Representational structure figures into many theories of psychological processing—of the *dynamics* of our psychological states. Theories of psychological processing aim to explain how we come to be in certain psychological states under certain conditions. What is the normal temporal ordering of the states in question? Which states causally depend on which others? What principles govern why one state causes or is caused by some other state or states? For example, how do we go from registering a layout of light intensities on the retina to perceiving an arrangement of surfaces in three-dimensional space? How do we go from seeing something to recognizing it as, for instance, a car? On the now common approach laid out by David Marr (1982), many psychological theories describe and explain causal relations between psychological states in terms of processes that build up, break down, traverse through, compare, or otherwise manipulate the part-whole structures of those states. Roughly speaking, the causal processes that are sensitive to representational structure are *computations*.

In order to give an account of how a structure-sensitive process produces or operates on a psychological state, one must give an account of the structure of that psychological state.

Psychologists typically assume that object recognition, for example, involves comparing a perceptual representation to a remembered representation. Almost all theories of object recognition suppose that this comparison process operates on the structures of the online perceptual representation and the stored representation. These theories treat the similarity between the two representations as a function of how the structures of those representations are related. For example, one very simple comparison process might match the perceptual representation with the stored representation just in case the former shared sufficiently many parts, combined in enough of the same ways, as the stored representation. The task of a theory of object recognition is to specify, on the one hand, what kinds of part-whole structures characterize the perceptual and memory representations that are input to the recognition process, and, on the other hand, how the recognition process goes about comparing these structures. Some models of object recognition suppose that the representations in question have as primitive constituents representations of simple three-dimensional shapes (Marr and Nishihara, 1978; Biederman, 1987). Others suppose that the primitive constituents are representations of surfaces (Nakayama et al., 1995; Leek et al., 2009). Recognition processes may compare structures by virtue of how identical they are in their parts and their structural relationships, or whether the representations can be related by some designated type of mathematical transformation, or what have you (see Sutherland, 1968; Marr and Nishihara, 1978; Ullman, 1996).

Understanding the structure of a representation can be central to understanding why it expresses a certain temporal or causal signature. For example, the psychologist Stephen Palmer (1977) conducted a classic experiment in which he looked at what sorts of structures must be involved in *part verification*, where perceivers must determine whether one set of elements is

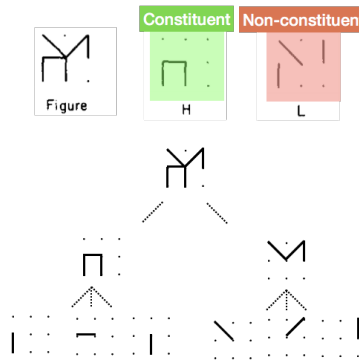


Figure 1.1: Top left: A “figure” comprising six line segments. Top Middle and Right: Two “part probes,” each comprising three line segments. Palmer argued that the part probe labelled “H” corresponds to a unified constituent of one’s representation of the figure, whereas the part probe labelled “L” does not. Bottom: The tree illustrates a hypothesis about how one’s perceptual representation of the whole figure decomposes into representations of component parts, which themselves decompose into representations of individual line segments. (Adapted from Palmer, 1977, p. 445).

contained within another set. For stimuli, Palmer used combinations of line segments, as illustrated in Figure 1.1. “Figures” were arranged from six component line segments and “parts probes” were arranged from three component line segments. Palmer then presented subjects with a series of displays consisting of a figure and a part, side by side. For each display, subjects were instructed to indicate, by pressing a button corresponding to either “Yes” or “No,” whether the displayed part was present in the adjacent figure. Subjects’ reaction times were categorically faster for some part probes than for others.

In order to explain this categorical difference in reaction times for different kinds of part probes, Palmer hypothesized that one’s representation of a figure has as constituents representations of subsets of line segments in the figure, and that these constituents are themselves structured from representations of the individual line segments. Palmer hypothesized that in order to detect the presence or absence of a set of line segments within the whole figure,

the visual system performs a matching procedure: beginning with the whole representation, this procedure checks whether any of the immediate constituents of the representation are representations of the relevant set of line segments. If not, the procedure checks the basic constituents for representations of the individual line segments in question. Subjects are faster to detect the presence, in the figure, of some subsets of line segments (for example, the subset labelled “H” in Figure 1.1), Palmer argued, because those subsets are represented by unified constituents of our representation of the whole figure. Subjects are much slower to detect the presence of other subsets of line segments (for example, the subset labelled “L” in Figure 1.1) because the representations of the individual line segments are not unified into a common constituent. In these cases, the matching procedure has to check multiple branches, at multiple levels, of the representation of the whole figure. As a result, the matching procedure takes much longer to determine whether representations of the line segments in question are in fact constituents of the representation of the whole figure.

Notice that Palmer’s studies directly address not just the *perceived* structure of the scene (how we take the figure to be organized from component line segments), but also the structure of the perceptual representation itself, which is the object of psychological processing. Palmer’s account serves to illustrate how hypotheses about how perceptual states are structured can join together with hypotheses about perceptual processes (for example, matching procedures) in order to explain the causal and temporal relationships between psychological states. It is unclear how to explain Palmer’s results properly except by positing some sort of structure-sensitive matching procedure, together with a hypothesis about the structures of the perceptual representations over which the matching procedure operated.

Palmer's part verification paradigm continues to be used in order to investigate the structures of our perceptual representations. For example, Edwin Leek et al. (2005) employed the logic of the part verification paradigm to compare Irving Biederman's theory that perceptual representations of object shapes have as constituents representations of three-dimensional part shapes and the theory that instead the constituents of object-shape representations are representations of two-dimensional surface shapes. Leek et al.'s stimuli consisted of outlines (in fact, line drawings) of novel three-dimensional objects. Potential "parts" were defined, in different conditions, as open contours, closed contours, contours that corresponded to volumetric parts, and contours that corresponded to arrangements of two-dimensional surfaces. They found that reaction times were faster for contours corresponding to volumetric parts than for open or closed contours, but they also found an equivalent advantage for contours corresponding to two-dimensional surfaces. They concluded that it is possible that representations of two-dimensional surface shapes, rather than three-dimensional volumes, be constituents of representations of object shapes.

Gordon Baylis and Jon Driver (1993) likewise formulated hypotheses about how our perceptual representations of space are structured in order to explain the causal and temporal dynamics of those states. They presented subjects with ambiguous stimuli in which one might either perceive one central object or two flanking objects (similar to Edgar Rubin's famous face-vase display). They indicated a point on each dividing contour in the stimulus, and asked subjects to indicate with the press of a button which point was vertically lower. Various manipulations were used to bias subjects to have a one-figure or two-figure percept. The stimuli were set up so that on different trials one or the other of the flanks, when perceived as a figure,

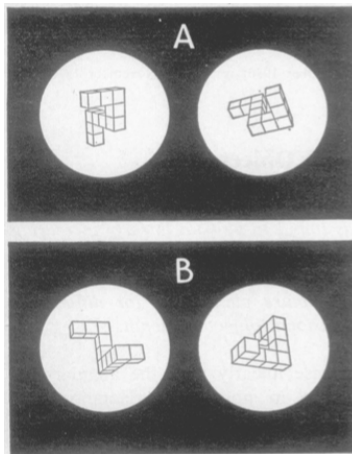


Figure 1.2: (A) A “same” pair, which differs by an 80° rotation in the picture plane; (B) A “same” pair, which differs by an 80° rotation in depth. (From Shepard and Metzler, 1971, p. 702).

would be perceived as vertically higher or lower than the other. Baylis and Driver found that the speed and accuracy of responses under the two-object conditions were significantly worse than under the one-object conditions, even when the physical stimuli were identical.⁵

Outside of perception, structural hypotheses have been at the center of understanding the processes by which mental images—as when you close your eyes and imagine a dachshund on a desk—are generated and manipulated. Mental imagery has long been at the center of debates over the structure or, as it is often also called, the “format” of psychological representations (see Kosslyn, 1980; Pylyshyn, 2002). In “mental rotation” tasks, subjects are asked whether two figures are rotated versions of each other (for example, Figure 1.2). When the

⁵Barenholtz and Feldman (2003) showed a similar effect for spatial comparisons within and between parts of an object: within-part performance was qualitatively faster than between-part performance, even when the absolute spatial relationships between the points in question were identical. This suggests that representations of elements that are perceived as within a part are structurally closer than representations of elements that are perceived as between distinct parts. Timothy McNamara proposed an analogous hypothesis that our cognitive representations of spatial regions are stored “in different branches of a graph-theoretic tree. The mental representation is organized such that increasingly more detailed spatial knowledge is given at lower and lower levels of the hierarchy” (McNamara, 1986, p. 90). (See also Duncan, 1984; Behrmann et al., 1998; Scholl, 2001).

two figures are rotated versions of each other, the time it takes subjects to match the figures is directly proportional to the angle by which the one figure is rotated from the other (Shepard and Metzler, 1971). It seems that subjects are performing some functional analogue of “rotating” the figures in their imagination. In “mental scanning” tasks, subjects are presented with an image or scene—for example a map or an outline of a boat, or objects in a box—which is then taken away. Subjects are asked to first imagine one point on the image or scene, such as the bow of the boat. They are then asked to “scan” across the remembered image or scene and report on whether some element is present, for example a rudder, or if the element has some feature. When the element is present, subjects take longer to respond the farther away the elements were from each other in the presented image (Kosslyn, 1973; Kosslyn et al., 1978; Pinker, 1980).

Mental rotation and scanning studies show that in certain kinds of tasks the time it takes to respond to the task is a function of the represented spatial relation (either angle of rotation or distance) between the key, task-relevant stimuli. Stephen Kosslyn and many others have taken these kinds of data to suggest that the representations involved in mental rotation and mental scanning tasks have as constituents both representations of the task-relevant stimuli that the subjects are asked about and representations of spatially intermediate stimuli, and that rotation and scanning operations must traverse through the representation’s structure from one constituent to the other. In fact, Kosslyn suggests that one can define a distance metric between parts of a mental image, and that this distance metric correlates to the represented spatial distance between the things represented by those parts.

Representational structure plays a central role in explanations of the causal and tempo-

ral relations among psychological states. Many of our perceptual capacities seem to involve some sort of structure-sensitive processing. Without an account of representational structure, accounts of structure-sensitive processing, such as the matching procedure posited by Palmer, cannot get off the ground. Moreover, different hypotheses about structure have different consequences about perceptual processing.⁶

Hypotheses about representational structure are integral to processing explanations of perceptual responses. But it would be a mistake to suppose that hypotheses about representational structure only function to support stories about the causal and temporal relations among psychological states. Psychologists commonly pose hypotheses about representational structure that do not, in themselves, make specific commitments about the order or timing of perceptual processes (for explicit statements to this effect, see Palmer, 1977, p. 449; van der Helm and Leeuwenberg, 1996, p. 430; Feldman, 1999b, p. 138–9).⁷

Consider the hypothesis that perceptual representations of the shapes of objects (humans, trees, and butterflies) are structured from perceptual representations of the three-dimensional shapes of the parts of those objects (limbs, boughs, and wings). This hypothesis places some constraints on the nature of the processes that produce and operate on the hypothesized struc-

⁶Indeed, Fodor (1975) argued forcefully, focusing on cognitive and psycholinguistic capacities, that the most successful theories of psychological processing posit structure-sensitive processes (computations) and that such theories are committed, on the face of it, to structured psychological representations over which such processes can operate. If we take the success of these theories at face value, then, we should infer that it is likely that there really are structured psychological representations.

⁷The analogous point is well-established in generative linguistics, where there is little assumption that the syntactic structure of an expression determines a particular parsing strategy for recovering that syntactic structure from an input string, although of course the available parsing strategies are constrained by the expression's syntactic structure, and *vice versa* (see, for example, Chomsky, 1965, p. 9; Berwick and Weinberg, 1983). Clowes (1971, p. 79–81) is likewise emphatic that structural theories are not committed to specific processing theories in the context of machine vision.

ture. Not every causal process could get from the sensory information at the retina to a state with that kind of structure, and not every causal process could make use of states in virtue of their structure. And yet there are many ways to produce such a structure. The representations of each part could be generated one after another, serially, and then combined. The representations of each part could, alternatively, all be generated simultaneously, in parallel. The complex representation of the whole object could be generated after the representations of the parts have been produced, or the complex representation could be formed all together along with its parts. There are also many things one can do with such structures. They may be stored in memory in various ways, matched to representations, used to anticipate new information, and so on. And each of these procedures may be carried in many different ways. Hypotheses about structural relations among representations vastly underdetermine hypotheses about the causal and temporal relations in which those representations stand.

What do structural theories explain, and how can we test them, when they are not figuring into, or accompanied by, processing accounts of the causal relations between psychological states? What is the explanatory and empirical import of structural theories that abstract from or generalize over processing theories? There are very general properties shared by the processes that produce or operate on a given kind of structure. The processes that can produce or operate on a given type of structure may be characterized by a particular kind of computational complexity. Part of the relative generality of theories of representational structure might consist in the fact that theories of representational structure can function to explain the presence or absence of these more general processing properties. That a perceptual representation has a given type of structure may explain in general why that representation is particularly

resource-intensive to produce, even abstracting the nature of the specific processes involved. Conversely, evidence about the complexity of a psychological process may offer evidence for or against a structural theory, even if one cannot give an account of the algorithm being computed (Tsotsos, 1993). However, psychologists do not merely, or even primarily, exploit structural theories in order these abstract computational properties. In the next section I will describe how theories of representational structure can function to explain what I will call the *demographics* of psychological states. One way that structural theories can have empirical and explanatory value independently of processing theories is by explaining and predicting the demographics of psychological states.

1.3 The Demographics of Psychological States

Theories of how psychological states are structured, together with accounts of what the primitive constituents of those states are, are well-suited to explaining what I call the *range* and *distribution* of psychological representations. The range of representations in a system is the set of representations that can nomologically occur in that system. The distribution of representations concerns how representations in that system and their properties can, cannot, or must nomologically co-occur. Where the system is a psychological one, the range and distribution of that system concern which representations can *psychologically* occur (range), and which sets of representations can, cannot, or must psychologically co-occur (distribution).

Structural theories help explain the range and distribution of representations in a system by appealing to *structural possibility*. Some combinations of psychological states will be

structurally possible while others will not. Indeed, an important part of specifying a mode of combination by which perceptual states can be parts of a more complex perceptual state is to specify the constraints on which perceptual states can and cannot be combined in that way. Constraints on combination can serve to explain why certain perceptual states can or cannot occur, or why they can, cannot, or must co-occur with other perceptual states. An ideal and unified account of the structures of perceptual representations would specify a set of unstructured, primitive representations along with a set of modes of combination, such that (a) all the representations that occur in that system are structurally possible combinations of primitives, (b) the structural impossibility of certain combinations of perceptual states explains why their corresponding complex perceptual states cannot occur in the perceptual system, and (c) what a representation's constituents are and how that representation can combine with others into complexes explains law-like patterns in how that representation co-occurs with other representations.⁸

There are modals, here, in both the *explananda* and in the *explanans*. One wants to explain why certain perceptual states are psychologically possible or not, for instance. The modal claims about the range and distribution of perceptual states are inferred on the basis of experimental data. The modality at stake in describing the range and distribution of psychological states is one of *psychological possibility* (and *necessity*, and so on). A task of psychology

⁸The psychologist's appeal to the structure of perceptual representations in order to explain the range and distribution of perceptual representations is analogous to the linguist's appeal to syntax to explain the possible expressions (or not) of a language and the distribution of those expressions within larger complexes. The explanatory role of psychological representational structure also recalls the traditional conception of logical form as delimiting the scope of and containment relations between what can possibly be thought.

One might say that the structural constraints that explain range and distribution correspond to "formation" laws, or a "grammar," whereas processing theories explain causal relations by appeal to "transition" laws, or laws of computation and inference.

is to explain in virtue of what certain psychological states are possible or not. Of course, one also wants to explain why certain psychological states *do* or *do not* occur, for example in experimental settings. But very often one wants to explain more law-like generalizations as well as make counterfactual assertions about what psychological states *could* occur, given the appropriate circumstances. I will not attempt to clarify how data about actual psychological performance can justify modal claims about what we are psychologically capable of. This is an instance of a fundamental question—maybe one of *the* fundamental questions—in the philosophy of science more broadly: how do observations of actual instances justify nomological claims?

One may explain psychological possibility by appealing to a narrower, more well-defined modality. One may explain the range and distribution of our psychological states by appealing to, among other things: (a) which psychological states are possible given the physiological limits of our sensory organs; (b) which psychological states are possible given limitations on resources such as attention, memory, and time; and (c) which psychological states are possible given the structural constraints on the combinability of states. Since my interest here is in the nature of representational structure, I will focus on at how hypotheses about *structural possibility* may be invoked to explain more general nomological claims about about what perceptual states (and what co-occurrences of perceptual states) are and are not possible. As we will see, structural possibility is not reducible to input-relative or resource-relative possibility.

I will describe a set of hypotheses that I take to be positing representational structure in order to explain the range and distribution of perceptual states. Unfortunately, psychologists

commonly equivocate between how we perceive things to be structured (*represented* structure) and how our perceptual representations of those things are structured (*representational* structure). It is not always clear which a psychologist intends to describe. For example, studies of “perceptual organization” are often ambiguous about whether they concern how we perceive a scene to be organized or how our perceptual representation of the scene is itself organized.

Crucially, accounts of representational structure, in contrast to accounts of represented structure, figure into causal psychological explanations. It is easy to see that accounts of structure-sensitive processing concern the representational structure of perceptual states, since it is the structure of the perceptual state itself that is supposed to stand in causal relations to other psychological states. But one must also be talking about the structure of perceptual states themselves if one is going to explain the range and distribution of those perceptual states. In such cases, it is the structure of the psychological state itself that putatively explains why it can be formed and how it can and cannot be combined with other psychological states. The causal efficacy of representational structure is central to explaining facts about the range and distribution of representations in a system. If psychological processes were not sensitive to the hypothesized structure, we would have no reason to think that that structure has anything to do with what representations can be produced and how they co-occur. So we can disambiguate a psychological hypothesis as a hypothesis about representational structure, rather than represented structure, insofar as the hypothesis aims to give a causal explanation either of psychological processing or of the range and distribution of representations in the system.

1.3.1 Range

In explaining the *range* of representations in a system, we want to explain why some representations are in principle possible elements of the system while other representations are not. For example, a theory of the syntax should explain why competent speakers find “Ada eats the cheese” an acceptable piece of English but not “The eat Ada cheese.” A syntactic theory provides an idealization of such acceptability judgments, according to which “Alma eats the cheese” is a possible sentence of English whereas “The eat Ada cheese” is not. The theory may then explain why the former sentence, unlike the latter, is a possible sentence of English, by showing how it is structurally possible according to combinatorial rules that seem to generalize. Importantly, the range of a representational system might not be exhaustively explained by the structural constraints governing that system. There is a sense in which the following sentence is outside the range of competent English speakers: “The man who said that a cat that the dog that the boy owns chased killed the rat is a liar.” Plausibly, the sentence is structurally possible according to the rules of English syntax. It seems likely that psychological limitations on short-term memory, rather than structural constraints on combinability, exclude the sentence from the range of a speaker’s competence (Miller and Isard, 1964).

The cardinality of representations in a system is one example of an important feature of that system’s range. We can call an organism’s representational capacities *productive* if the organism has finite psychological resources (memory, processing time, and so on) yet it is capable of producing any of infinitely many—or, in any case, very, very many—distinct representations. There is some reason to think that perception is productive, insofar as we are able

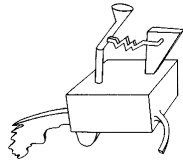


Figure 1.3: A “nonsense” shape that one has likely never seen before and that one likely cannot classify under any familiar categories. (From Biederman, 1987, p. 116).

to perceptually represent an enormous number of novel, distinct shapes and scenes. For example, you are able to represent the shape of the “nonsense” shape in Figure 1.3, despite never having seen it before and not being able to classify it under some familiar category. Fodor and others have argued that if an organism’s representational capacities are productive, the best explanation is that the organism is using a representational system in which representations can be combined in a finite set of ways from a finite set of primitive representations (see, for example Fodor, 1987). For example, it may well be that the visual system is capable of producing an enormous number of shape representations because the visual system is able to combine in determinate ways a basic repertoire of shape representations. So, the general claim that representations in a system have structure can explain how a limited organism can employ or instantiate a representational system with an expansive range. The abductive inference has roughly the following form:

Productivity/cardinality That a basic set of representations can combine into complex representations can explain how a system with finite resources can have infinitely many—or at least very, very many—different kinds of representations.

What is sometimes called “systematicity” is also a feature of range. As Fodor explains it, a representational capacity is *systematic* if “the ability to produce/understand some of the

sentences is intrinsically connected to the ability to produce/understand many of the others. . . . You don't, for example, find native speakers who know how to say in English that John loves Mary but don't know how to say in English that Mary loves John" (Fodor, 1987, p. 149). These sorts of considerations about cardinality and the presence of systematicity do not say much about the *specific* structures that those representations have. Other kinds of facts about the range of a representational system can be useful in formulating specific structural hypotheses.

There has been substantial debate over how exactly to characterize the property of systematicity (see, for example Cummins, 1996; Aydede, 1997; Johnson, 2004). McLaughlin (2009) suggests that there are really a variety of specific "systematicity laws," which assert that "ceteris paribus, a cognizer has a certain thought ability if and only if the cognizer has a certain other thought ability" (McLaughlin, 2009, p. 253). We do more to characterize the actual structures of representations in a system by specifying these particular laws than we do by noting that the representations are systematic *simpliciter*.

For example, there is good reason to think that one can visually represent as of a *red circle beside a blue square*, if and only if, all things equal, one can also visually represent as of a *blue circle beside a red square*. One explanation for this systematicity law is that these representation share common constituents—representations of *redness*, *blueness*, *circularity*, and *squareness*—combined in different ways (see Treisman and Gelade, 1980; Tacca, 2011). A related systematicity law is that one can represent as of a *red circle* if and only if, all things equal, one can represent as of a *blue square*, *redness*, and *circularity*. This law can be explained by the hypothesis that the representations as of a *red circle* and as of a *blue square* have different constituents, but a common mode of combination—what one might call *feature*

binding. These explanations of systematicity have the following form:

Systematicity That a representational system includes one kind of representation if and only if it includes another kind of representation can be explained by positing that those representations share either constituents or a mode of combination.

Structural theories are often invoked to explain more general facts about which kinds of representations are and are not included in a representational system. Recall the Palmer (1977)'s part verification experiments. Palmer found that subjects were categorically faster at detecting the presence in a figure of some sub-patterns than they were at detecting certain other sub-patterns. To explain this difference in reaction times, Palmer suggested that some sub-patterns received unified representations while others did not. The detection of sub-patterns in a figure involves a visual process that operates over one's representation of a whole figure and proceeds, in top-down fashion, to check whether any of the constituents of that representation represent the elements of the sub-pattern. The procedure will terminate more quickly if the relevant sub-pattern is represented by a unified constituent than if the elements of the sub-pattern are represented by a variety of disconnected representations. So far, we have an account of what the parts of a given representation are, but not an account of why some patterns receive unified constituents while others do not.

Palmer noted that subjects' were faster to verify the presence of patterns that were intuitively very "good" forms, or good Gestalts, while they were slower with patterns that were intuitively "bad" forms. Palmer had independently solicited reports on what patterns struck subjects as "good" or "bad." While this sort of task may seem wildly ambiguous, subjects'

responses were well predicted by the standard perceptual grouping principles suggested by early Gestalt psychologists. According to those principles, the perceived “goodness” of a set of elements in a scene depends on, among other things, whether the elements are sufficiently close together, whether they are connected, whether they are continuous, and whether they form a closed boundary. Palmer showed that a weighted sum of factors like proximity, connectedness, continuity, and closure predicted what we see as good forms. Putting it all together, Palmer argued that it is just the patterns that were particularly good, according to this metric, which receive unified representations. In other words, representations of different line segments will combine into a representation of a group of line segments only if those line segments, as represented, pass a threshold for being sufficiently proximal, connected, and so on. As Palmer put it:

The part probes with high goodness values will tend to be encoded as single SUs [structural units] at a high level in the hierarchical network representing the figure. Part probes with low goodness values will tend to be encoded as multiple SUs at lower levels in the representation of the figure, since the segments will be embedded in more than one higher-order SU. If SUs are processed as integral units, then a “good” probe will be verified more quickly and/or accurately than a “bad” probe because the good probe will require less SUs to be verified than the bad probe. (Palmer, 1977, p. 456)

Palmer’s theory purports to explain not just dynamic features of one’s perception of patterns, but also the range of one’s perception of patterns. First, subjects are faster at verifying

the presence of some sub-patterns because those sub-patterns receive unified constituents. Secondly, representations of elements can combine into complex representations of patterns only if those elements, as represented, are sufficiently proximal, connected, and so on. Versions of this latter point have also been made about contour grouping (for example Geisler and Super, 2000; Elder and Goldberg, 2002). Interestingly, if Palmer's interpretation is right, the way we perceive the world to be organized is in fact reflected by the way our perceptual state itself is structured: when we perceive a sub-pattern as a "good" or natural part of a whole figure, our representation of that sub-pattern is a constituent of our representation of the whole figure.

Recall also Baylis and Driver's study of spatial perception. They found that subjects were faster and more accurate at perceptually judging the spatial relationships between points on the same object than when judging the relationships between points on different objects—even if the spatial relationships were in fact the same. Baylis and Driver interpreted this comparative advantage for the one-object conditions as evidence that there exists a common integrated representation of the object that has as constituents representations as of points on the same object while the representations as of the points on different objects can only be constituents of separate representations as of distinct objects. This suggests a constraint on how perceptual representations of elements can combine into representations of how those elements are spatially related: representations of the points on an object can immediately combine with each other, while representations of the points on different objects cannot. In order to represent the relations between the points on different objects, the representations of those points must be constituents of representations that are themselves combinable. As with the Palmer study, perceived structure (one object versus two objects) seemed to correspond to

the structure of the perceptual state itself, with perceived parthood corresponding to constraints on structural combination.

The following abductive principle seems to be at play in these structural theories, which I will state with maximum generality:

Exclusion That a type of representation is not psychologically possible can be explained by positing that the representation is not structurally possible given the modes of combination and primitive representations that characterize the relevant psychological system.

For example, that it is not psychologically possible to represent a certain set of line segments as a group can be explained by positing that such a representation cannot be combined from the relevant representations of line segments.

There do not seem to be compelling non-structural explanations of the studies mentioned above. One could not explain the psychological impossibility of the relevant perceptual states by appeal to a lack of appropriate sensory stimuli. In both studies, subjects were required to respond to aspects of the very same physical stimulus. In Palmer's study, subjects repeatedly had to detect the presence of one or another set of line segments (depending on the trial) within the very same figure. The sensory stimulus produced by a given figure could in principle have led to different ways of grouping the parts of the figure. The fact that subjects grouped some line segments but not others cannot be explained by physical aspects of the stimulus alone. The same point applies to Baylis and Driver's study. They employed ambiguous stimuli, manipulating a subject's perceptual set so as to bias her to perceive the two points as lying either on the same object or on separate objects. The sensory stimulus was physically identical in

each of these conditions. The physical stimulus itself cannot account for why spatial representations could combine in different ways in the different conditions.

Further, it is not clear how general resource limitations could explain the results of either study unless they presupposed the hypotheses about representational structure discussed above. Baylis and Driver offered their structural theory as part of an explanation for why attention is facilitated when comparing points on a common object than when comparing points across objects. Attention, they suggested, must operate according to the structures of representations. It is easier to attend to a set of elements if one's representations of those elements are immediate constituents of a common object representation. It is unclear how to explain the relevant constraints on attention without supposing that attention operates on the basis of representations with certain kinds of structure. A similar point applies to attempts to explain Palmer's data in terms of attentional limitations. It is easier to attend to "good" groups than "bad" ones *because* it is easier for the visual system to delegate attention on the basis of a unified structural constituent of a representation than on the basis of multiple disconnected constituents (see Vecera, 1994; Kimchi, 2009).

It will be instructive to consider a case in which there has been a debate over the proper explanation of the range of a perceptual capacity. Daniel Schacter and his colleagues employed priming paradigms to investigate what kinds of complex shapes can be perceptually represented (Schacter et al., 1991). Priming studies measure how subjects respond to a target display as a function of being briefly exposed to a prior display. Perceptual priming involves the automatic, unconscious recall of a representation in perceptual memory in order to assist in the generation of a later representation. If the representation of the current stimulus is of

the same kind as the stored representation of an earlier stimulus, then the stored representation will benefit, or “prime,” the processing of the new representation, with the result that the reaction based on the second token is faster and/or more accurate (see Tulving and Schacter, 1990; Ochsner et al., 1994).⁹ Schacter et al. investigated whether exposure to the contours of an “impossible object” would facilitate one’s ability to respond to overall properties of that same object upon a later exposure. Subjects were exposed to a view of an impossible object and then, later to the same view of that object. One task was to determine whether the perceived object was, on the whole, facing left or right (see Figure 1.4).¹⁰ They found that exposure to the contours of impossible objects did not facilitate later responses about the properties of that same, whole configuration of contours, whereas such facilitation was found with exposure to the contours of possible objects. They argued that the reason we do not find a priming effect is that we do not form perceptual representations of the whole impossible object in the first place, whereas we do form perceptual representations of the whole possible object. They predicted that we do represent consistent parts of impossible objects, and so one would find priming on tasks that required attending to the parts of the impossible object rather than to the whole thing.

⁹Perceptual priming differs from associative priming, which occurs with semantic memory. *Perceptual memory* is distinct from *semantic memory*, and these can exhibit different empirical signatures. For example, perceptual memory storage, recall, and degradation are typically faster than semantic memory storage, recall, and degradation. Perceptual memory is largely encapsulated from influence by long-term memory and conceptual associations. The two types of systems are also doubly dissociable: individuals may have damaged perceptual memory capacities but intact semantic memory capacities and *vice versa*.

¹⁰The stimuli in these experiments consisted of line drawings of objects. The use of line drawings in vision science experiments can complicate the interpretation of results. The important thing to note here is that Schacter et al. were inquiring into subjects’ abilities to perceptually represent three-dimensional objects and not into their abilities to represent pictures or to represent the depictive content of pictures. On the intended interpretation of these experiments, a line drawing is a convenient substitute for a genuinely three-dimensional stimulus containing nothing but the contours of an object as cues to the object’s shape, excluding other information about shape such as binocular and textural depth cues.

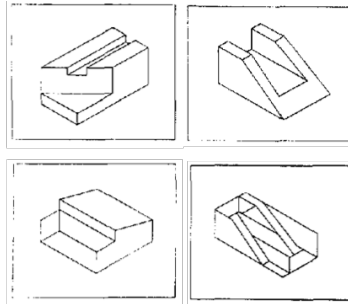


Figure 1.4: Examples of stimuli used in (Schacter et al., 1991). The objects in the left column, on the whole, face to the left while those in the right column face to the right. The upper row contains geometrically possible objects and the bottom row contains geometrically impossible objects. (Modified from Schacter et al., 1991, p. 4).

Why do we find occurrences of perceptual representations of possible objects but not occurrences of perceptual representations of impossible objects? Schacter et al. proposed that perceptual representations of object shapes are *structural descriptions*—they represent the shapes of the components of the objects as well as the structural relations between those objects (see also Marr and Nishihara, 1978; Biederman, 1987). They suggest that the system that produces such representations

cannot settle in on a single global interpretation of an impossible object, precisely because there is no globally consistent interpretation of the structure of such an object. The structural description system can, however, compute a globally consistent interpretation of a possible object, and it is this representation that we assume provides a basis for priming. (Schacter et al., 1991, p. 16)

Schacter et al.’s language here is a bit misleading. Typically, to say that a representation is *consistent* is to say that it represents a logically possible situation—it has a model. Of course, there are representational systems that contain representations of logically impossible situa-

tions. “This is a square and it is not a square” is an inconsistent but legitimate sentence of English. We are able to form thoughts about logically impossible situations, as when we realize that some thesis leads to a particular contradiction. The point extends to notions of arithmetical and geometrical consistency and possibility. “ $2 + 2 = 5$ ” is a well-formed arithmetical expression, though it represents an arithmetical impossibility. It does not follow, *in general*, that because a situation is impossible, there can be no representation that purports to represent that situation. In other words, it is not generally true that only consistent representations can be tokened. Schacter et al.’s point seems to be that in the special case of the human visual system there cannot be perceptual states that represent as of geometrically impossible objects because such states would not themselves be structurally possible—they literally cannot be put together.

Schacter et al.’s structural hypothesis is formulated in terms of what a computational system, which is dedicated to producing representations of objects, can and cannot output. This reflects the point, made above, that representational structure must ultimately be the product of causal psychological processes. Any constraints on the combinability of representations must be realized by some computational system that outputs the combinable representations but not the uncombinable ones. At the same time, the structural hypothesis substantially underdetermines exactly what that computational system looks like. One can think of the structural hypothesis as indicating the presence and function of the relevant “structural description system,” without specifying how the system is computationally realized.

Carrasco and Seamon (1996) challenged Schacter et al.’s studies, arguing that their experiments failed to establish that subjects cannot form unified perceptual representations of

whole impossible objects. Carrasco and Seamon noted that the stimuli that Schacter et al. employed for impossible objects were more complex than the stimuli employed for possible objects. Carrasco and Seamon controlled for the complexity of the stimuli, using different measures of complexity—including objective measures concerning, for example, the number of angles, as well as subjective measures acquired from subjects' reports about the complexity of the stimuli. They found that less complex impossible figures did support priming to the same extent as possible figures of equivalent complexity. To explain why more complex impossible figures do not support priming, Carrasco and Seamon consider the hypothesis that

the structural description system is capable of computing global structural descriptions of both possible and impossible objects, providing those objects are not structurally highly complex. . . . For structurally complex objects, such as the extreme[ly complex] impossible objects of the present study, lack of priming may be due to insufficient time and resources available for encoding. (Carrasco and Seamon, 1996, p. 350)

Carrasco and Seamon take for granted that the representations of objects are structured. However, if they are right, then Schacter et al.'s findings are less the result of structural constraints than of resource constraints. Carrasco and Seamon suggest that the reason certain perceptual representations as of whole impossible objects are not possible has little to do with their *structural impossibility*. It may well be that the constituents of those representations are in principle combinable, but that they cannot be formed due to resource bottlenecks. According to this hypothesis, it is limitations on more peripheral, general-purpose mechanisms and

resources that accounts for our inability in certain cases to represent as of whole impossible objects. The structural description system itself may have been in principle capable of forming those representations, but for these external constraints. The authors do not seem to count constraints on time and memory as *structural constraints*.

If one's inability to produce a specific perceptual representation can be traced entirely to limitations in sensory inputs, resources, or systems that are not primarily dedicated to realizing that perceptual capacity, then there is not sufficient basis for supposing that the state is *structurally impossible*. Structural constraints are not reducible to these other psychological limitations. Instead, structural constraints on which representations of a given kind are structurally possible—for example, which object representations are structurally possible—are realized by intrinsic constraints on the processes that are dedicated to producing that kind of representation. Carrasco and Seamon's response to Schacter et al. demonstrates how one might empirically distinguish between structural constraints on psychological states and peripheral constraints on the general resources and systems that are employed in processing those states.

1.3.2 Distribution

The range of representations concerns which representations can or cannot occur at all. Inquiries into the range of psychological representations often dovetail with inquiries into the distribution of representations. The distribution of representations concerns how representations can, cannot, or must *co-occur*. For example, why is it that anytime the expression “Elisbet” occurs in a well-formed phrase of English (for example, “Elisbet frowns”), we can sub-

stitute “Alma” for “Elisabet” (for example, “Alma frowns”) and obtain another well-formed English phrase? Other distributional facts concern how representations *must* co-occur. Why, in general, does the occurrence of a sentence necessitate occurrences of a verb phrase and of a noun phrase? Linguists explain these distributional patterns in terms of constraints on syntactic structure. “Alma” and “Elisabet” are of the same structural category, in the sense that anytime one can occur as a constituent of a well-formed expression, the other could replace it while preserving the well-formedness of the whole. The reason that every time a sentence is uttered so are a verb phrase and a noun phrase is that sentences must be made up of verb phrases and noun phrases.

Turning again to perception, many studies into how we perceive scenes to be organized make claims about representational structure on the basis of distributional patterns. One type of distributional pattern is present when the occurrence of kind of representation necessitates the occurrence of other representations. For example, I cannot form a representation of the tabletop in front of me as brown and rectangular without forming a representation of the tabletop’s brownness and a representation of its rectangularity. A representation of color and shape as co-instantiated in an object necessitates individual representations of that color and shape (Treisman and Gelade, 1980; Tacca, 2011). A natural explanation for this is that the former representation has the latter representations as constituent parts, and a complex representation cannot be tokened without its parts. Notice that the distributional pattern here is a contingent feature of how our perceptual system works. One could design a representational system with a primitive representation for brown, rectangular things, which could be tokened without also tokening individual representations of brownness and rectangularity.

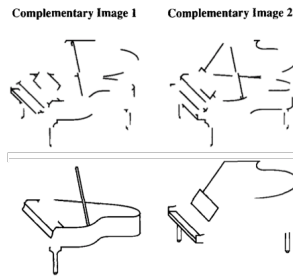


Figure 1.5: Top row: feature-deleted outlines. Bottom row: component-deleted outlines. (Biederman and Cooper, 1991, Adapted from).

Patterns of necessitation are at the center of Irving Biederman and Eric Cooper’s (1991) experiments on our perception of three-dimensional shapes. In their experiments, they presented subjects with outlines (in fact, line drawings) of familiar objects, where intersections or junctions of contours in those outlines were deleted (Figure 1.5). Contour junctions, as well as sharp concave curves, are important cues for perceptually identifying the three-dimensional shapes of an object’s parts (see Hoffman and Richards, 1984; Enns and Rensink, 1991; Bhatt and Bertin, 2001). Biederman and Cooper conducted two experiments to examine priming effects of junction-deleted outlines. For our purposes, the key question was how one junction-deleted outline would prime the same outline in which the complementary set of junctions was removed (for example, the two columns in Figure 1.5).

In their first experiment they used what they called “feature-deleted” outlines, in which all object parts had at least one junction cue in the outline (for example, the top row in Figure 1.5). In their second experiment they used what they called “component-deleted” outlines, in which some object parts had no junction cues (for example, the bottom row in Figure 1.5). Biederman and Cooper found priming for complementary feature-deleted outlines. They argued that it was unlikely that subjects were forming representations of the deleted junctions

on the basis of contour completion mechanisms. So they concluded that the same shape could be represented without representing arbitrary junctions on that object. Biederman and Cooper found no priming for complementary component-deleted outlines. So they concluded that subjects could not represent the same shape without representing the same set of component shapes. Biederman and Cooper took this to confirm the hypothesis that representations of object shapes have as constituents representations of the three-dimensional shapes of the object's parts and representations of the relations between those parts.

The general pattern of inference here is something like this:

Necessitation That the occurrence of one perceptual representation necessitates the occurrence of another can be explained by positing that the latter is a constituent of the former.

It is often in principle possible to construct non-structural, merely causal explanations for distributional patterns. Suppose we accept the interpretation that visual representations of three-dimensional shapes necessitate individual visual representations of the shapes of their three-dimensional parts. One might try to explain this pattern by proposing that representations of the parts *cause* representations of the whole shape. This clearly would not explain the pattern of necessitation, however, since it is open that shape representations could be caused in the absence of representations of their component parts. One would have to say that representations of parts are necessary for causing representations of whole shapes. However, this is not yet a developed alternative to Biederman and Cooper's account. On their account, representations of parts will be necessary for causing representations of whole shapes, because

the former are literally constituents of the latter. As a result, the parts must be caused either before or together with the whole representation.

In order to explain the pattern of necessitation without appeal to constituency, but *merely* by appeal to causal relations, one would have to say either that representations of parts are *the only possible cause* of shape representations, or that shape representations themselves *necessarily cause* part representations. The latter option seems to be a nonstarter. In the first place, one would have to explain why one could not interrupt the causal chain, blocking the formation of the part representations just as the shape representation is formed. In the second place, and more importantly, Biederman and Cooper showed that by intervening on the formation of part representations, one can interrupt the formation of the shape representation. So, if anything the causal order is from part representations to shape representations, not *vice versa*.

On the other hand, the claim that part representations are *the only possible cause* of shape representations seems unlikely on its face. One would want to know why there are no other possible causes of shape representations. Indeed, Biederman's theory that part representations are constituents of shape representations could help to explain why you *must* cause the former in order to cause the latter. Although this causal story is not a strict implication of Biederman's theory, it would be a natural addition that theory. However, the mere causal account cannot appeal to the structural story in order to explain why shape representations can only be caused by part representations.

In fact, the main competitors to Biederman and Cooper's theory are other structural theo-

ries. These theories share the same underlying methodological assumptions about how to find out about structure. For example, Leek and his colleagues (2009) gathered additional distributional data to argue that there was no reason to favor Biederman's theory over an alternative hypothesis. They found that object outlines containing cues for three-dimensional part shapes showed no priming advantage over displays containing only cues for visible two-dimensional surface shapes. They inferred that while there is evidence from priming that representations of object shapes required representations of visible surfaces, there is no priming evidence that representations of those shapes also required representations of three-dimensional part shapes. So they concluded that there was no evidence favoring the hypothesis that complex shape percepts consisted of three-dimensional part shape percepts over the hypothesis they defended in an earlier paper, that "the basic units of shape description consist of 2D edge-bounded primitives that are used to approximate object surface shapes" (Leek et al., 2005, p. 679).¹¹

Another kind of distributional pattern concerns how properties of a representation can vary independently of each other. Consider Kent Stevens' (1983) discussion of the nature of our visual representations of surface orientation. Stevens distinguished between the hypothesis that our perceptual representations of *surface orientation* consist of representations of *slant* (the angle between the line of sight and the normal of the surface) and *tilt* (the angle or direction in which the surface's normal would project to the image plane), and the hypothesis that

¹¹Both Leek and Biederman are primarily concerned with the nature of the perceptual representations that are inputs to object recognition processes. Biederman and Cooper provide some reason to believe that their priming experiments are targeting *perceptual* representations and are not making use of additional conceptual information. In short, where Biederman and Cooper did find priming between complementary junction-deleted outlines, they did not find priming between junction-deleted outlines of different shapes of the same category (for example, between an outline of a grand piano and an outline of a saloon piano). In addition, while Leek and Biederman suppose they are examining representations that are inputs to recognitional processes, these experiments do not rely on recognitional tasks or processes, but rather on perceptual priming.

our perceptual representations of surface orientation consist of representations of the surface's depth gradient along the x -axis of the image plane (designated " p ") and the surface's depth gradient along the y -axis of the image plane (designated " q ").

These are different hypotheses about the constituents of our perceptual representations of surface orientation. Presumably, on neither hypothesis could one represent surface orientation without tokening both of the putative constituents. But these hypotheses differ substantially in how they predict and explain the distribution of precision or noise in our percepts of surface orientation. For example, Stevens refers to evidence that our perception of surface orientation is more precise along the dimension of tilt than along the dimension of slant. This is not easily explained by the p, q hypothesis. Because values of p and q do not independently entail values for slant and tilt, it would be very surprising to find that representations of p and q regularly co-occurred in just the right way to produce a systematic difference in noise along the tilt versus slant dimensions. By contrast, if our representations of surface orientation are made up of constituent representations of slant and tilt, it would be far less surprising to find systematic differences in the occurrence of noise along the slant versus tilt dimensions. Stevens argues that the hypothesis that surface orientation percepts are made up of distinct representations of slant and tilt, which can independently differ in their noisiness, best explains the distribution of noise in our perception of surface orientation.

The abductive inference pattern here seems to be:

Independence That perceptual representations of a given kind have independently varying properties can be explained by positing that those properties are derived each from sep-

arate kinds of constituents.

E.J. Green and Jake Quilty-Dunn (2017) follow a similar pattern of inference in a very different case. They reject the hypothesis that representations of color are necessarily also representations of shape and orientation—that is, they reject the hypothesis that color, shape, and orientation are not represented by separate constituents. They appeal to evidence that how much one forgets one of these features is independent of how much one forgets another of the features (see Fougne and Alvarez, 2011). If the fidelity of color representations can deteriorate independently of the fidelity of orientation representations, it is reasonable to infer that color and orientation are represented by separate constituents.

In fact, these distributional patterns of *independent variation* can be presented in a slightly different way, which will be more familiar from the study of language. These studies essentially show that the representation of a particular tilt or color can be substituted for another representation—either of a different tilt or color, or merely a noisier representation of the same tilt or color—in the context of the same representation of some other property, such as slant or shape, respectively. Representations of slant and representations of tilt make up their own structural kinds, or equivalence classes, where members of one can be interchanged in the context of a given member of the other.

So, we may posit separate types of constituents in order to capture the fact that constituents of those types can vary independently of each other. On the other hand, there are cases where we have antecedent reason to think that a representation has distinct kinds of constituents, yet there are structural *dependencies* among those constituents. Consider Figure 1.6.

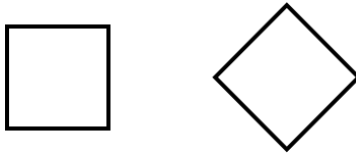


Figure 1.6: Our representations of shape and orientation are structurally interdependent.

You likely see a square on the left and a diamond on the right. Yet the left figure could also be perceived as a tilted diamond and the right figure could be perceived as a tilted square. In general, one tends to see a square as an upright square if its sides are parallel to the edges of the display, whereas we tend to see it as an upright diamond if its corners point to the edges of the display (see Rock, 1973). Noting this, Humphreys and Quinlan (1988) argued that perceptual representations as of squares are structurally distinct from perceptual representations as of diamonds. They further argued that this structural difference depended on how one represented the orientation of the figure (see also Humphreys, 1983).

Humphreys and Quinlan found that percepts as of squares did not prime percepts as of diamonds. Priming studies measure how subjects respond to a target display as a function of their being briefly exposed to a prior display (the probe). The operational assumption is that the perceptual representation of the probe display is stored in an “implicit memory” buffer that can be accessed by perceptual processes (see Tulving and Schacter, 1990; Ochsner et al., 1994). If the target display induces the same type of representation as the probe, then the perceptual system can use the stored representation of the probe in order to facilitate the response to the target (e.g. speeding up one’s reaction time for identifying the target). In this case, we say that there is a “priming effect,” or that the representation of the initial display “primes” the representation of the target display. Priming between representations is evidence

for similarity in representations, whereas the lack of priming is evidence that the representations are different in kind. While priming studies address the nature of representations in implicit memory, we can make a defeasible assumption that the structure of a representation stored in implicit perceptual memory retains some (but perhaps not not all) of the original structure of the memorized perceptual representation.¹² Since representations of squares and diamonds did not prime each other, Humphreys and Quinlan argued that those representations are structurally distinct.

The lack of priming between representations as of squares and representations as of diamonds was not due merely to a difference in perceived orientation. Representations of an unambiguous figure (an isosceles triangle) at different orientations did prime each other. The important thing is that whatever orientation we attribute to the square must bisect the side of the square and not its corners. That we can perceive a shape as that shape at different orientations suggests that whatever state represents as of that shape (a square, a diamond, an equilateral triangle, or an isosceles triangle) does not have a representation as of orientation (relative to the scene or to the viewer) as a constituent. This suggests that representations of orientation (relative to a scene or viewer) are not constituents of shape representations, but rather combine with shape representations.

We have reason, then, to think that there are representations of oriented shapes which have as constituents a representation of shape and a representation of orientation. The constituent representations of shape and orientation seem to be interdependent. The “reference

¹²Perceptual priming must be distinguished from conceptual priming. These have different empirical signatures and are neurophysiologically dissociable. Unlike perceptual priming, conceptual priming can rely extensively on associative rather than structural connections between representations.

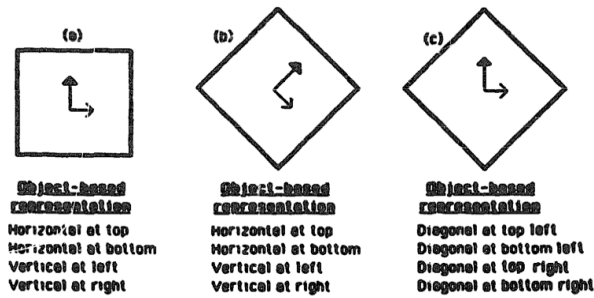


Figure 1.7: An illustration of a reference frame-based shape representation. (From Hinton and Parsons, 1988, p. 2).

frame theory” of shape perception purports to explain this interdependence as a structural interdependence (see Humphreys, 1983; Palmer, 1989). According to this theory—or really, this family of theories—we perceive shapes in terms of how the parts of those shapes are located and related with respect to an object-based reference frame. We can roughly think of shape representations as specifying coordinates or qualitative locations (such as *top*) for parts the shape as well as relations like *above*, *parallel*, and so on (Figure 1.7). To represent a square is, among other things, to represent a horizontal line at the top of the figure, a horizontal line at the bottom, and vertical lines to the side. An orientation representation anchors the *up* of a shape representation’s reference frame to some direction in the scene (or some direction based on the viewer). What cannot happen, structurally, is that the represented orientation of a figure relative to the scene does not line up with the shape’s own intrinsic up-down axis. A representation of the left figure in Figure 1.6 as *upright* cannot combine with a representation of that figure as a diamond, whose intrinsic *up-down* axis runs through the corners of the figure. The tokening of a particular orientation representation will select for an appropriate type of shape representation, and *vice versa*.

In its most abstract form, the reference frame theory offers the following sort of explanation:

Dependence That occurrences of one kind of perceptual representation depend on occurrences of another kind of non-constituent representation can be explained by positing a structural constraint on how states of those kinds can and cannot combine.

Notice that while the reference frame theory offers an account of the structural interdependence between representations of shape and representations of orientation, the theory does not commit to any specific account of how this interdependence is causally realized. Structural theories can be used to explain patterns of necessitation, independence, and interdependence among perceptual states, while abstracting from the particulars of how those states are causally and temporally related.

1.3.3 Summary

To summarize, structural hypotheses can function to explain the range and distribution of psychological representations. The following abstract, approximate patterns of abductive inference seem to be employed throughout the empirical research:

- (1) *Range*: Which representations can and cannot occur in the system.

Productivity/cardinality That a basic set of representations can combine into complex representations can explain how a system with finite resources can have infinitely many—or at least very, very many—different kinds of representations.

(For example, representations of novel shapes (Biederman, 1987; Fodor, 1975).)

Systematicity That the inclusion of one kind of representation implies the inclusion of another kind of representation can be explained by positing that those representations share either a constituent or a mode of combination.

(For example, representations of feature conjunctions (Treisman and Gelade, 1980; Tacca, 2011).)

Exclusion That a type of representation is not psychologically possible can be explained by positing that the representation is not structurally possible given the modes of combination and primitive representations that characterize the relevant psychological system.

(For example, limits on perceptual grouping (Palmer, 1977); hierarchical spatial representation (Baylis and Driver, 1993); representations of impossible objects (Schacter et al., 1991; Carrasco and Seamon, 1996).)

(2) *Distribution*: Which representations can, cannot, or must co-occur in the system.

Necessitation That the occurrence of one perceptual representation necessitates the occurrence of another can be explained by positing that the latter is a constituent of the former.

(For example, representations of feature conjunctions necessitate representations of individual features (Treisman and Gelade, 1980); representations of three-dimensional object shapes necessitate representations of three-dimensional parts (Biederman and Cooper, 1991) or else two-dimensional surfaces (Leek et al., 2009).)

Independence (Substitutivity) That perceptual representations of a given kind have

independently varying properties can be explained by positing that those properties are derived each from separate kinds of constituents.

(For example, independent noise in representation of slant and tilt (Stevens, 1983); independent forgetting of color and shape features (Fougnie and Alvarez, 2011; Green and Quilty-Dunn, 2017).)

Dependence That occurrences of one kind of perceptual representation depend on oc-

currences of another kind of non-constituent representation can be explained by positing a structural constraint on how states of those kinds can and cannot combine.

(For example, interdependence of representations of shape and orientation (Rock, 1973; Humphreys, 1983; Palmer, 1989).)

Structural hypotheses will typically predict and explain a cluster of features concerning range and distribution, while also being apt to figure into specific accounts of perceptual processing. For instance, consider the hypothesis that representations of the three-dimensional shapes of objects must have as constituents representations of the three-dimensional shapes of the parts of those objects and representations of the relations among those parts. This hypothesis predicts that one cannot represent the three-dimensional shape of an object without representing the three-dimensional shapes of its parts and the relations among those parts (*necessitation*). The hypothesis also predicts that representations of the shapes of object parts can vary independently from the representations of those parts' spatial relations (*independ-*

dence/substitutivity, see Arguin and Saumier, 2004). And the hypothesis, together with certain assumptions about visual processing, predicts that representations as of the parts of an object will be processed more easily together than the representations as of the parts of different objects (cf. Baylis and Driver, 1993; Barenholtz and Feldman, 2003). The attribution of representational structure can support explanations of a variety of different phenomena that might otherwise appear to be disunified.

A structural theory explains the psychological possibility (or impossibility) of a perceptual state—or of the co-occurrence of several perceptual states—by reference to its *structural* possibility (or impossibility). Structural explanations of range and distribution are not reducible to, and in fact may trade off with, explanations that appeal to physiological limits on sensory inputs or to general resource limitations. Structural constraints on a certain type of representation are realized not by these peripheral limitations but by intrinsic features of the processes dedicated to producing those representations. Peripheral explanations of range and distribution often go hand in hand with structural accounts. Structural explanations of range or distribution may sometimes compete with hypotheses that appeal solely to causal relations among states. It is always a case-by-case, empirical matter which of these alternative explanations—structural, peripheral, or hybrid—is strongest.

1.4 The Generality of Structural Theories

I will now argue that structural theories can abstract away from the details of causal processes when explaining and predicting features of range and distribution.¹³ Processing theories function to explain and predict the temporal and causal relationships between psychological states. The range and distribution of representations in a system underdetermine the *explananda* of processing theories. In the next section, I will argue that the range and distribution of representations in a system are a source of evidence for structural theories that is independent of evidence for particular processing theories.

Most of the structural hypotheses that we have considered do not specify the processes that either produce or operate on the hypothesized structures. For example, Biederman hypothesized that representations of the three-dimensional shapes of whole objects are structured from constituent representations of the three-dimensional shapes of object parts. This structural hypothesis does not specify the causal and temporal relations among object and part representations or between those representations and other psychological states. The complex representation might be formed through serial processes or parallel processes. The complex representation might come after, and causally depend on, the formation of the constituent representations. Another possibility is that the complex might be formed together all at once on the basis of a common antecedent (for example, representations of the overall symmetry and coarse-grained metric properties of the object) and that only subsequently can

¹³It is reasonable to suppose, though I have not argued for it here, that the architecture of a given system—the system's basic operations and the ways they compose with each other—determine the structures of the representations in that system. But the converse is not true: different kinds of processes can operate over the same structures, yielding the same range and distribution of representations.

constituent representations be “broken off” and operated upon independently of each other (Navon, 1977; Kimchi, 2015). Another possibility is that the perceptual system first tokens an abstract schema or prediction specifying very generally what sort of complex representation is to be produced and then fills in the particular constituents in whatever order possible, with room for feedback and revision as the structure is fleshed out (see Love et al., 1999; Yuille and Kersten, 2006). Or some combination of these possibilities might occur. Structurally basic or primitive states, and constituents in general, need not be causal antecedents of the complexes of which they are parts.

Moreover, while Biederman hypothesizes that object recognition involves processes that compare structures of this sort, he does not specify what those comparison processes actually are. Moreover, Biederman’s experiments into the existence of these structures did not just depend on recognitional processes. Biederman also looked into the extent to which these structures support priming. One might look at any number of different systems and types of processes when gathering evidence for a structural hypothesis. Structural theories like this are substantially open-ended and noncommittal about what the specific causal processes that produce and operate on the complex representation look like.

Structural theories can be noncommittal about the specifics of processing when their aim is to explain range and distribution, because these *explananda* are more general than the *explananda* of processing theories. The range and distribution of psychological states are compatible with many systems of temporal and causal relations between states. Suppose that occurrences of representations of the three-dimensional shapes of objects necessitate occurrences of representations of the three-dimensional shapes of object parts. It could be that the

latter occur first and cause the former. But it might also be that both are always caused at the same time. Further, the distributional pattern is equally consistent with the claim that the representations of the shapes of the parts might be generated serially, one after the other, or in parallel. The distributional pattern is also consistent with different claims about how representations of whole object shapes feed into recognitional processes as well as other types of systems. Features of the range and distribution of representations in a system underdetermine the causal and temporal relations between those representations. If a structural hypothesis functions to predict and explain features of range and distribution, then we can see how that hypothesis might prescind from specific processing hypotheses.

It may help to clarify the sense in which structural theories generalize over processing theories by situating this claim with respect to Marr (1982)'s well-known distinction between *computational* theories and *algorithmic* theories. Roughly, a computational-level theory describes at a high-level of abstraction how content about some property can be derived from information registered by sensory receptors, assuming that certain constraints are satisfied in the perceiver's environment (see also Kitcher, 1988). For example, a computational-level theory may explain how content about a surface's *size* can be derived from content about its *distance* and information about its *visual angle*, assuming that the environment satisfies certain geometric constraints (Kaufman et al., 2006). An algorithmic-level theory describes how the input registrations of sensory information are structured, how the output representations of distal content are structured, and by what effective, step-by-step procedures the perceptual system transitions from the input registrations to the output representations. Of course, these different levels of description are really an idealization. In reality, we can theorize about

psychological systems at many levels of abstraction (see Palmer and Kimchi, 1986).

Since, as I have argued, structural theories abstract from detailed accounts of processing, it would make sense to locate them at the computational level of description. But notice that computational-level theories may make some general commitments about causal processing relations between states—for example, that the representation of size *depends* on the representation of distance together with the registration of visual angle. Structural theories can prescind from even the very abstract causal claims that are common at the computational level. A theory about how three-dimensional shape representations are structured may not be committed to any specific theory of what sensory cues or prior representations the shape representations causally depend on. So there is a sense in which structural theories can be even more general than many computational-level theories—or at least more general than the fragments of those theories that are concerned with specifying causal dependencies.

At the same time, we can see from the discussion in Section 2 that structural theories purport to describe the same structures on which concrete, causal procedures at some algorithmic level operate. So, Marr is right to associate theories of representational structure with the algorithmic level at which detailed accounts of processing are given. As I see it, structural theories crosscut the computational and algorithmic levels. Structural theories generalize across both fine- and coarse-grained accounts of the causal processing relations between states, while specifying structures that enter into those causal processes.

1.5 An Independent Channel of Evidence

I emphasize the relative generality of structural theories, because it has more often been supposed that the empirical validity of any given structural theory is hostage to whatever processing theory accompanies it. The psychologist John Anderson, for example, argued that “Any claim for a particular representation[al structure] is impossible to evaluate unless one specifies the processes that will operate on this representation. Arguments for or against a particular representation are only valid assuming a particular set of processes” (Anderson, 1978, p. 250). Anderson attempted to provide a general proof that for any distinct structural theories, one can produce accompanying processing theories that will predict all the same data. He supposed that one cannot confirm or disconfirm a structural theory, short of having evidence about the specific processes operating on those structures. Anderson was pessimistic that we could ever come up with such empirical evidence. However, even among those who are more optimistic, the optimism usually rests on the assumption that we can find evidence for a particular structural hypothesis because it is possible to find evidence in favor of an associated processing theory (see Pylyshyn, 1979, 1984; Block, 1983; Camp, 2007; Johnson, 2015).

The shared assumption among the pessimists and optimists is that it is only by way of confirming a processing theory that one can confirm a structural theory. However, this assumption fails to recognize that structural accounts figure into explanations and predictions of range and distribution, not simply explanations of the temporal and causal relations between states. Structural theories can be confirmed or disconfirmed by data about range and distribution, independently of any concrete theories about how the representations in ques-

tion are processed. One consequence is that there are actually more channels of confirmation for theories of structure-sensitive processing than have usually been recognized in theoretical discussions.

Anderson argues that any computational process defined over one kind of structure, or grammar, can be simulated in both its input-output behavior and its major steps by a computational process defined over a distinct grammar. He makes the point vivid by presenting two different hypotheses about the structure of mental images and the processes operating over them. Anderson claims that the two hypotheses are indistinguishable on the basis of the behavioral data. He designs his illustration around a letter rotation task (see Cooper and Shepard, 1973). In this task, a figure is presented to the subject, and the subject must decide whether the figure is or is a rotation of a letter of the alphabet. One must explain why performance on the match-trials takes longer with a greater angle of rotation. On one structural account, performance on this task depends on processing representations with array-like structures: a representation of the whole letter has the structure of an array whose cells contain representations of segments of the letter, where the structural relations between cells is isomorphic to the spatial relations between the represented letter segments (see Chapter 3). On the array account, matching requires going through some analogue of incremental matrix rotation.

On the alternative “propositional” account, matching relies on representations with structures similar to the formulas in first-order predicate logic: a predicate, denoting some relation or feature (such as the vector angle), is applied to some representations that refer to elements of the figure (such as its component line segments or its intrinsic axis of orientation). On this propositional account, matching requires going through incremental substitutions for the

vector angle that is attributed to the figure's axis of orientation. Both of these theories seem capable of explaining the behavioral data that reaction time on match trials is a linear function of angle of rotation. But the theories offer very different accounts of the representational structures of the perceptual states involved. Anderson claims that there is no way to empirically decide between the accounts, short of determining what the relevant processes are.

Many of the responses to Anderson's argument attempt to show that one can gather empirical evidence about how some representations are processed without having any independent evidence about how the representations being processed are structured (see Pylyshyn, 1979, 1984; Block, 1983; Camp, 2007; Johnson, 2015). I want to point out that the opposite is true as well. One can gather empirical evidence about how psychological representations are structured without having any independent evidence about how those representations in particular are processed. In particular, structural theories can predict, and so be confirmed or disconfirmed by, patterns in the range and distribution of representations in a system without specifying the processes that operate on the hypothesized structures.

On the array account, a representation's width and height are not represented by separate constituents. Rather, the very same set of filled cells in the array that compose a representation of width and of height. By contrast, if the spatial representations involved in mental imagery have the structures of formulas in classical predicate logic, with separate predicates denoting width and height, then there ought to be well-formed representations as of width that need not co-occur with representations of height, and *vice versa*.¹⁴ If we can find width-only or height-

¹⁴Of course, there are many other structural accounts one could offer. Different structural accounts can overlap in their predictions on specific tests. For example, it could be that the representations involved in mental imagery do have subject-predicate structure, but have a more restrictive syntax than the formulas of classical

only representations in the relevant system, by whatever experimental means, then the array account seems to have made a bad prediction about the distribution of representations. If we do not find such width-only representations, then the propositional account should give some additional, independently confirmable story about why the perceptual system never does form representations that, according to the account, are well-formed. Indeed, Fournie and Alvarez (2011) looked for, and failed to find, width-only representations in their study of memory degradation.¹⁵

The set of ways one can find out whether there are such width-only representations is open-ended, and carries no particular commitments about how the relevant representations are generated or employed. No well-defined class of input-output behaviors exhausts all the evidence for some feature of the range or distribution of representations. One could look into, among other things, matching performance (as with Palmer's studies), patterns of priming (as with Biederman and Cooper (1991)'s study), or patterns of memory degradation (as with Fournie and Alvarez (2011)'s study). It is true that in each of these paradigms, there will be some assumptions about how representations are processed—how visual matching is performed, or how representations are stored in memory. Yet structural theories need not be committed to any particular one of these processing assumptions. Moreover, many of those processing as-

predicate logic, according to which a well-formed representation must contain both a predicate denoting width and a predicate denoting height. This account would make the same prediction as the array account here: there should be no experimental trace in normal subjects of length representations that fail to co-occur with orientation representations, and *vice versa*. While different accounts may concur in their predictions on a given test, they can be expected to differ in their abilities to predict and explain other phenomena.

¹⁵One must be careful in interpreting negative results like this. Nevertheless, there is a burden on the propositional account to offer an independently motivated explanation for why there was no evidence for width-only representations. In Chapter 3, I offer some positive arguments that a core class of perceptual representations are structured like matrices or arrays.

sumptions concern general purpose systems—systems of visual matching or memory—which are not dedicated exclusively to either generating or operating on the kinds of representations being investigated.

While structural theories need not make specific commitments about processing, processing theories often do make commitments about structure. These embedded structural hypotheses make an independent empirical and explanatory contribution to the processing theories into which they figure. Processing theories can be evaluated partly on the basis of whether their structural predictions about range and distribution are borne out. Suppose we have a theory that outlines an algorithm for generating object-shape representations that consist in a set of representations that locate contours and surface features at various distances relative to the perceiver—something like a segmented Marrian “2.5D sketch” (Marr, 1982; Jackendoff, 1987). If object-shape representations are nothing but sets of representations of features at locations, then representations as of impossible object shapes should be well-formed. A representation as of an impossible object shape would consist of representations of the object’s features at locations. It will turn out that it is geometrically impossible for all of those constituent feature representations to be jointly accurate. Yet it should nevertheless be possible, on this “feature-placing” account, to form the representation. We can check the structural hypothesis by looking for evidence that there are such representations—or, if there are not, whether this is due to facts about processing power or implementation.

If a structural hypothesis is embedded in a process theory—if a theory presents a psychological process as operating on a certain kind of representational structure—then we should be able to develop predictions about the range and distribution of representations in the given

system. It is because structural hypotheses function to explain and predict features of range and distribution, and these features underdetermine processing relations, that structural hypotheses can abstract over and provide an independent source of evidence for the processing theories into which they figure.

1.6 Psychosyntax and Psychosemantics

We perceptually represent the world as structured. In addition, much of contemporary perceptual psychology takes perceptual representations *themselves* to be structured from—or made of—other perceptual representations. I have not really tried to give a positive argument that perceptual representations *are* structured. Instead, I have tried to clarify what is involved in saying that a perceptual state, or in general any psychological state, has representational structure. I have taken a non-reductive approach, focusing on the explanatory role of attributing structure to perceptual states. I expect that by better understanding that explanatory role, we will be in a better position to judge whether we should be realists about structure.

Representational structure, or syntax, is a kind of part-whole structure that is paradigmatically had by representations. Representational structure plays a central role in explaining the causal relations between psychological states. Appeals to representational structure can also abstract from accounts of the causal relations between psychological states. Structural theories can explain more general phenomena concerning what psychological states we can be in (the *range* of psychological representations) and the patterns in how psychological states can or must co-occur (the *distribution* of psychological representations). As a result, structural theo-

ries can be confirmed or disconfirmed independently of processing theories. This last point is good news for theories of structure-sensitive processing. One can justify the processing claims and the structural claims of those theories using multiple, relatively independent channels of evidence. The widespread view that structural theories are empirically hostage to processing theories seems to stem from neglecting the role that structural theories play in explaining the demographics, and not just the dynamics, of psychological states.

What is the relationship between how our perceptual states are structured and how they represent the world to be? This question is at the heart of the next chapters. I will defend a principled constraint on this relationship: what a complex perceptual state represents depends exhaustively on what its constituents are, what they represent, and how they are combined (Chapter 2). I will also argue for a more substantive relationship: the structural forms of perceptual states carry substantive content about the spatial arrangements of the things that those states represent (Chapter 3). Finally, I will argue that the distinctively perspectival character of perception is best explained by the relationship between the structures of perceptual states and their contents (Chapter 4). For now, I want to point out what appears to be a pervasive dependency of our descriptions of structure on our descriptions of content.

Notice that in every case that we have discussed, the specification of representational structure has made some reference to what is represented. For example, Schacter et al. (1991) specifies a semantic constraint on modes of combination for object representations: the resulting representation must be as of a geometrically possible object. Biederman (1987) and Leek et al. (2005) define constituents of representations of object shapes by the contents of those constituents: the constituents are either representations of volumetric shapes or representa-

tions of two-dimensional surfaces. Palmer (1977) argued that representations of line segments can combine into a representation as of a group if those line segments, as represented, are related—including their proximity, closedness, connectedness, and so on.

In each of these cases structure is specified in part by reference to what the states involved purport to represent. Complex representations are partly picked out by their content (for example, that it is a representation as of an object shape). Constituents are partly picked out by their contents (for example, that they are representations as of three-dimensional shapes or that they are representations as of two-dimensional surfaces). Modes of combination are specified in terms of either the contents of the to-be-combined constituents (i.e. that they be representations as of “similar” contours) or the content of the resulting combined representation (that it be a representation as of a geometrically possible object).

The representational structure of a psychological state is central to that state’s causal dispositions. Philosophers have sometimes proposed that, since psychology is primarily interested in describing the causal relations among psychological states and how they give rise to behavior, psychologists could describe psychological states in terms of their structural properties without reference to their representational content. Reference to the contents of those states, this view suggests, is inessential to the scientific project of psychology, or it is merely part of an informal “gloss” of what those states are (see, for example Stich, 1983; Egan, 2014). However, reference to representational content seems to be indispensable to developing and testing empirical theories about the kinds of representational structure discussed here (see also Crane, 1990; Burge, 2010a, pp. 95–6).¹⁶

¹⁶A somewhat separate matter is whether representational structure depends *metaphysically* on semantic prop-

An early claim of the generative syntax program in linguistics was that one need not describe the semantic properties of natural language expressions in order to describe their syntactic categories or the principles governing their combinations with other expressions. Noam Chomsky, for example, argued that semantic theory should not inform or place constraints on syntactic theory in the study of natural languages (see Chomsky, 1957). It may well be, though it is controversial, that in the case of natural language one could identify the limits and patterns of occurrence and co-occurrence of expressions without reference to the semantics of those expressions.

The case of perceptual psychology seems to be very different. It is not clear how we could specify the range and distribution of such representations without making any reference to their contents. We are not in a position to type-identify perceptual states with neural states or with specific causal roles. So, it is natural to describe the structures of psychological representations at least partly in terms of their contents (see also Crane, 1990; Burge, 2010a, pp. 95–6). There seems to be an insight hiding in the confusion, often found in the empirical literature, between represented structure and representational structure: in all of the cases we have discussed, the structure of a perceptual state is specified by reference to aspects of the represented structure of the world.

erties. Philosophers and psychologists have at various times and for a variety of reasons argued that representational structure must be metaphysically independent of semantic properties. Even if the formulation and testing of structural descriptions depends in part on reference to semantic descriptions, the thought goes, it cannot be that the structure of a psychological state metaphysically depends on its content. This issue is deeply rooted in concerns about the mind-body problem, the metaphysics of causation, the interpretation of scientific explanations, the theory of computation, and G-d knows what else (for further discussion of this issue, see Fodor, 1983; Peacocke, 1994; Burge, 2007). My claims here are just about the centrality of content-ascriptions to the scientific project of specifying and testing theories of psychological representational structure.

CHAPTER 2

Composing Percepts

[O]ne of the things a scientific community acquires with a paradigm is a criterion for choosing problems that, while the paradigm is taken for granted, can be assumed to have solutions [E]xplicit statements of scientific law and about scientific concepts and theories . . . help to set puzzles and to limit acceptable solutions.

Thomas Kuhn

THE STRUCTURE OF SCIENTIFIC REVOLUTIONS

Perceptual states attribute properties such as shape, size, color, and texture to particulars in the world. In addition, perceptual states can be “made of” or structured from other constituent perceptual states. For example, it is plausible that my perceptual representation of an object, such as a leaf or a box, has as constituents my perceptual representations of the object’s shape and of its color, among other features (see, for example, Treisman, 1986). How does the structure of a perceptual state relate to what it represents?

Suppose I am looking at a yellow, rectangular tabletop. It may sound like a truism to say that I represent the tabletop as yellow and rectangular because my representation of the tabletop is made up of representations as of *yellowness* and *rectangularity* (and indeed repre-

sentations as of instances of those properties). Naturally, it also matters how the constituent representations are put together. Suppose I see the yellow, rectangular tabletop supported by a cylindrical red column. I do not simply perceive the table as yellow, red, rectangular, and cylindrical. I perceive the table as having a yellow, rectangular part on top of a red, cylindrical part. This is because in my representation of the table, my perceptual representation of yellowness is more immediately integrated with my perceptual representation of rectangularity whereas my perceptual representation of redness is more immediately integrated with my perceptual representation of cylindricality. So how the whole perceptual state represents the table depends on what its constituent perceptual states represent and how those constituent states are combined. This is an instance of a more general principle that perception is *semantically compositional*: what a perceptual state represents depends entirely on what its constituents represent and how those constituents are combined. In this case, the compositional nature of our perception of the table's shape and color seems to be so transparent it is barely worth pointing it out. However, it is not always transparent how perceptual representations are semantically composed from their parts.

A number of phenomena show that it is not always easy to say how, or even that, our percepts are semantically composed. First, perceptual representations are not always computationally pieced together, bit by bit, from a set of causally independent constituents. The computation of one perceptual state can interact with the computation of other non-constituent perceptual states. And though we tend to think that perceptual representations of complex patterns are made up of perceptual representations of the simpler elements of those patterns, the former may be computed either before or simultaneous with the latter. The causal order of

things does not always parallel the putative structural and semantic order of things.

Second, complex perceptual states often represent emergent features that are not or could not be represented by their constituent perceptual states. For example, we represent the bounding contours of surfaces as having features, such as *area* and *closure*, that constituent perceptual states could not attribute to the segments of those contours. This contrasts with the seemingly simple way in which the content of my representation of the tabletop intersects the contents of the constituent representations of yellowness and rectangularity.

Third, the way one represents an element of a scene regularly depends on one's representations of surrounding elements in the scene. For example, whether I represent a surface as lighter or darker depends on how I represent the surface's surrounding environment. So how my perceptual representation of the surface represents that surface seems to depend on more than the internal structure of that representation.

These three types of phenomena were emphasized repeatedly by the Gestalt psychologists (for example, Wertheimer, 1938b; Koffka, 1936; Köhler, 1947) and have continued to be the focus of much research in perceptual psychology. While there have been passing references in both philosophy and psychology to the idea that perception has a compositional semantics (for example, Fodor and Pylyshyn, 1988; Feldman, 1999a), there has been little to no discussion of what a compositional semantics for perception should look like in light of these phenomena.

I will present different ways in which we can treat these phenomena compositionally. One implication is that these phenomena are not really in tension with the claim that perception is compositional. But more central to my purposes is that there is a substantive and

fruitful research program to be had in formulating and testing compositional theories of perceptual phenomena like these. It is challenging, yet informative, to work out explicit, empirically plausible, compositional theories of perceptual phenomena. In subsequent chapters, I will argue that by attending to how perceptual representations are semantically composed, we can better address longstanding questions in the philosophy of mind about the nature of perception—including the question of how perception differs from thought, and the question of how to reconcile the objectivity of perception with its perspectival character.

In the next section, I will introduce the notion of compositionality in more detail and I will review some standard, general arguments for thinking that perception is compositional. In Sections 2.2, 2.3, and 2.4, I will show how we can give compositional accounts of the phenomena mentioned above. Such accounts can make important contributions to our understanding of perception. I will conclude, in Section 2.5, by recommending some of methodological advantages, for psychology, of explicitly articulating compositional semantic theories of perceptual phenomena.

2.1 Compositionality

I will say that a system of representations has a *compositional denotational semantics* (or, for short, is *compositional*) if and only if the denotational content of every representation in that system is a function of, and only of, the denotational content of its constituents and the way those constituents are structurally combined. By the *denotational content* of a representation,

I mean the object, property, or state of affairs that the representation purports to single out.¹ I will call a *theory* of a representational system “compositional” if the system is compositional under that theory’s description.

The language of propositional logic, for example, has a compositional semantics. This language contains elementary propositional expressions (“ p ”, “ q ”, ...), logical terms (“ \neg ”, “ \supset ”), and brackets (“[”, “]”) to indicate structure. Some expression $\ulcorner \alpha \urcorner$ is well-formed in this language if and only if α is either an elementary propositional term, or if $\ulcorner \alpha \urcorner$ has the form $\ulcorner \neg \beta \urcorner$, or if $\ulcorner \alpha \urcorner$ has the form $\ulcorner \beta \supset \gamma \urcorner$ (where $\ulcorner \beta \urcorner$ and $\ulcorner \gamma \urcorner$ are themselves well-formed expressions). The contents of expressions in this language are truth values (either *true* or *false*), or functions from truth values to truth values. Any expression of the form $\ulcorner \neg \beta \urcorner$ is true if and only if β is false. Any expression of the form $\ulcorner \beta \supset \gamma \urcorner$ is false if and only if $\ulcorner \beta \urcorner$ is true and $\ulcorner \gamma \urcorner$ is false. So the value of every complex expression is a function of, and only of, the values of its immediate constituents and how those constituents are combined (for example, the order of the expressions flanking the term “ \supset ”).

That natural languages are compositional is one of the foundational commitments of contemporary semantic theorizing in linguistics (see, for example, Heim and Kratzer, 1998). On this picture, “Usain Bolt runs” means what it does because “Usain Bolt” denotes a particular person, “runs” indicates the property of *running*, and when the verb goes after the noun, the resulting sentence attributes the indicated property to the denoted individual. Natural languages contain many cases that are not so mundane. It is not easy to formulate a compositional

¹While the current discussion is restricted to the denotational semantics of perception, I intend the discussion to serve as a foundation for understanding how perceptual *modes of presentation* are compositional.

semantics that successfully predicts the meanings of expressions that contain, for instance, quantifiers, indexicals, or anaphora, and the sensitivity of word meanings to their surrounding syntactic contexts. While compositional theories of these phenomena may be available in principle, it is another matter to find a theory that is empirically adequate given our best standing theories of the syntax and semantics of perception (for further discussion, see Partee, 2004).²

Why should we think that perception is compositional? Indeed, why should we continue to hold to the principle that natural languages are compositional, given the existence of many phenomena that are difficult to handle compositionally? The arguments for compositionality in perception and in language roughly parallel each other. Since my interest here is perception, I will focus on arguments that perception is compositional.

For one, perceptual representations are *semantically productive*, in the sense that we could in principle token infinitely many—or in any case, very, very many—distinct perceptual representations, each with distinct contents. It is remarkable that we regularly encounter, and can accurately perceive, novel scenes that contain novel objects with novel shapes. In Chapter 1, I discussed structural explanations of the *productivity* of perceptual representations, which concerns the great (perhaps indefinite) number of distinct types of perceptual state. *Semantic*

²A number of people have argued that compositionality is formally vacuous (for example Zadrozny, 1994). Roughly, the claim is that for any set of representations one can set up or modify one's theory of syntax or semantics to satisfy compositionality. If these arguments are right, then we must be able, in principle, to give a compositional semantics for languages containing quantifiers, indexicals, anaphora, and so on. However, the structure and content of natural language representations, no less psychological representations, is an empirical matter. It is a substantive question whether a particular compositional theory of a representational system is empirically and explanatorily adequate. It may turn out that no compositional theory is empirically and explanatorily adequate. The project of formulating good compositional theories is a substantive, non-trivial research program in linguistics (for further discussion, see Westerståhl, 1998; Dever, 1999).

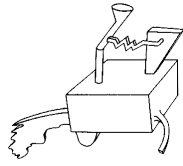


Figure 2.1: A “nonsense” shape that one has likely never seen before and that one likely cannot classify under any familiar categories. From Biederman, 1987, p. 116.

productivity consists of the fact that there are is a great (perhaps indefinite) number of *semantically* distinct types of perceptual state. You can, for instance, represent the strange shape of the object in Figure 2.1 although you have likely never seen that shape before. How can our perceptual capacities be so expansive and so accommodating of novelty? A structural explanation of semantic productivity will have to posit not just a structural hypothesis, that the perceptual states in question can be combined in a compact number of ways from a compact set of primitives, but also a semantic hypothesis that explains how structured representations can have novel contents that are not carried by their primitive parts.

Following Frege (1963, 1914), a number of people have argued that the best account we can give of how finitely limited creatures could token so very many semantically distinct representations is that these representations are semantically compositional and that they are structured in a finite set of ways from a finite set of primitive representations (see Fodor, 1987; Fodor and Pylyshyn, 1988). For example, the novelty of the shape of the object in Figure 2.1 has largely to do with the fact that simpler shapes are combined in a new way. Our capacity to represent the novel complex shape may depend on our ability to combine a finite set of capacities to represent the simpler, more common shapes and their spatial relations (see Marr and Nishihara, 1978; Biederman, 1987; Singh and Hoffman, 2001). The compositional story

is that our ability to perceptually represent indefinitely many distinct scenes, shapes, and so on, derives from our ability to combine representations of simpler, more common elements.

Perceptual representations are also *semantically systematic*. As I discussed in Chapter 1, *systematicity (simpliciter)* concerns the ways in which the inclusion of one kind of representation in a system nomologically necessitates the inclusion in that system of another kind of representation. Roughly, *semantic systematicity* obtains when those kinds of representations are not just nomologically related, but are semantically related in what they represent. Semantically related perceptual capacities seem to come as a package. If I can perceive as of a sphere on top of a cube, then I can perceive as of a cube on top of a sphere. It is something like a law of psychology that if I can perceive a scene as containing a blue square and a red triangle, then I can also perceive a scene as containing a red square and a blue triangle (see, for example, Treisman and Gelade, 1980). Many, from Frege to Fodor and on, have thought that the best account of why semantically related capacities should be linked is that they are semantically compositional. The capacities to perceptually represent a sphere on top of a cube and cube on top of sphere come as a package because those perceptual representations share common structural constituents and differ simply in the way those constituents are arranged. Those capacities are semantically related because those common structural constituents are also common semantic determinants.

Finally, compositionality seems to loom, though without much explicit attention or formal development, in the descriptive practices and explanatory aims of perceptual psychology. Typically when psychologists posit kinds of representational structure, they implicitly define that type of structure by providing what is in effect a compositional rule. For example, I

have made several references to examples of *feature binding*, as when a perceptual representation of a color combines with a perceptual representation of a shape. Psychologists typically define feature binding by specifying (a) that the perceptual states that can be feature bound are perceptual representations of properties (or indeed instances of those properties), such as shape and color, and (b) that the result of feature binding is a complex state that represents the properties denoted by its constituents as co-instantiated. Another way to combine perceptual states is to *configurally bind* them (Palmer, 1977; Geisler and Super, 2000). Configural binding consists in combining representations of individual items in a scene into a representation of a cohesive group or configuration. Effectively, in these cases, the mode of combination—feature binding or configural binding—is implicitly defined by way of a compositional rule that tells us how the content of the complex state derives from the contents of the combined states.³

Often it is not clear how to specify a type of representational structure in perception except implicitly through an associated compositional rule. Usually, neither neural descriptions of

³This practice tends toward something like *local* compositionality. A system of representations is locally compositional if and only if the contents of complex representations in the system are a function of, and only of, the contents of their *immediate constituents* and their mode of combination. In a locally compositional system, each condition on well-formed structural combination corresponds to a condition on the denotation of the resulting complex representation.

Following Montague (1970), Hodges (1998), and Janssen (2011), we can give a precise definition of local compositionality. The syntax and semantics of a representational system can each be described as an algebra. Each algebra contains a base set of elements (primitive constituents in one, primitive contents in another) and a set of operations or rules for specifying non-basic elements (structural rules of combination/semantic rules of composition). Each algebra determines a full universe of representations or contents. Let $\llbracket \cdot \rrbracket$ be the function that maps representations onto their contents. A representational system is locally compositional (relative to $\llbracket \cdot \rrbracket$) if and only if $\llbracket \cdot \rrbracket$ is homomorphism from the syntactic algebra to the semantic algebra. Where f is a structural mode of combination:

Local Compositionality A set of representations is locally compositional iff for every complex representation $f(\alpha_1, \dots, \alpha_n)$, there is a semantic rule g such that $\llbracket f(\alpha_1, \dots, \alpha_n) \rrbracket = g(\llbracket \alpha_1 \rrbracket, \dots, \llbracket \alpha_n \rrbracket)$.

representational structure nor functional descriptions that characterize representational structures entirely in causal terms, are ready to hand. It is not clear how to define feature binding, for instance, except by saying that a feature-bound state is a combination of representations of property instances, and that the feature-bound state represents the property instances denoted by its constituents as co-instantiated in a particular.

More fundamentally, a core aim of perceptual psychology is to find out the ways in which we perceptually represent the world and to explain how we are able to perceptually represent the world in those ways. In order to pursue this aim, psychologists have to join their characterizations of structure to their characterizations of semantics. We want to know not just how different perceptual states, each characterizing a certain part of the world, can combine, but also how their combinations provide a further characterization of the world. We would like to explain how some perceptual characterization of the world is possible in virtue of the combination of the ways some more basic perceptual representations characterize the world (cf. Burge, 2005, p. 194). This explanatory aim, together with constraints on how we can describe representational structure, lead perceptual psychology to proceed implicitly as if compositionality were the rule. To the extent that perceptual psychology is on the right track, then, we can infer that it is likely that perception *is* semantically compositional.

We have here three general arguments—from semantic productivity, from semantic systematicity, and from scientific practice—that perception is semantically compositional. These arguments parallel arguments that natural languages are compositional. There is, of course, room for debate about how similar the capacities of perception are to those of language. This debate requires us to attend not just to *whether* perception and thought are compositional,

but also to *how* they are compositional. Linguistics has largely moved past general arguments supporting compositionality and has evolved an extensive, directed effort to show how, specifically, natural languages are compositional, especially in the face of a variety of hard cases. This sort of research program demonstrates the interest, resilience, and fruitfulness of a commitment to compositionality. There is no comparable program in the empirical study of perception. While theories in perceptual psychology often tacitly suggest compositional theories, it is rare for psychologists to explicitly formulate and defend those theories. On the other hand, philosophers of perception have for the most part not moved beyond very general discussions about whether perception is compositional. What I want to do now is show that we can and should explicitly articulate compositional theories of perceptual phenomena.

2.2 The Causal Order and the Semantic Order

It will be instructive to begin by looking at how compositional semantic theories do and do not relate to theories of perceptual *processing*. Visual processing in humans is commonly characterized as hierarchical. On this picture, information about and representations of features of small bits of the world come to be progressively integrated into more complex representations of abstract features of larger regions of the world (for example, Grill-Spector and Malach, 2004). For example, the visual system first detects tiny bits of contours, texture, color, and so on, and progressively stitches them together into representations as of the shapes, sizes, and colors of larger chunks of the world, which themselves get stitched together until one has a representation of the whole scene. This picture of the visual system seems on first blush

to vindicate the claim that perception is compositional, for it looks as though the visual system literally starts by first producing semantically primitive representations and then, piece by piece, building up complex representations whose contents derive from the contents of the preceding representations.

Semantic compositionality is often presented as, or as concomitant with, a *computational* principle governing how we compute the contents of representations. It is easy to think of a compositional system of representations by analogy to a set of building blocks. On this picture, each primitive representation is like a basic building block. So long as they fit together (that is, so long as they are combinable at all), these blocks can be combined arbitrarily to produce complex structures with novel meanings. The building blocks are inert and do not exert any influence on each other. Moreover, one cannot produce a complex structure of blocks without building it up, piece by piece, from its basic parts. It can seem that one of the central advantages of having a compositional semantics is that it allows for this procedure of building up representations, piece by piece. Indeed, what we care about, one might think, is how perceptual states fit into the causal economy of our psychologies; if the semantic order does not parallel the causal order, then the semantic order is explanatorily superfluous. If the principle of compositionality has any explanatory value, one might think, it is because compositionality offers a model of how perceptual states are causally put together.

However, there are many reasons to think that a strictly hierarchical view of visual processing is not right. In the first place, it is common that the computation of one representation interacts with the computation of other non-constituent representations. Further, there is substantial evidence that the visual system computes representations of complex patterns at the

same time or even before it computes representations of the elements that make up those patterns. Just as the hierarchical picture seems to support the claim that perception is compositional, evidence that perceptual processing is not fully hierarchical might seem to undercut compositionality. If the building blocks model is essential to compositionality, then these facts about processing would seem to be incompatible with perception's being compositional. On the other hand, one might think that if compositionality does not imply a building blocks model of perception, it is not clear how a compositional theory could be illuminating. What value would a compositional semantics have if it the procedural generation of perceptual states does not parallel their semantic composition. I will indicate both how the causal facts mentioned above are compatible with compositionality and how having a compositional semantics might substantively add to our understanding of perceptual processes.

Perceptual processing can be highly interdependent. The computation of one representation may depend on the computation of another representation, even if the latter is not a constituent of the former. For example, the ability to discriminate and accurately perceive contour segments is better when those segments are parts of a continuous chain of contour segments (Polat and Sagi, 1993; Loffler, 2008), when they are parts of closed contours (Kovács and Julesz, 1993), and when they are perceived as parts of boundaries of three-dimensional figures (Williams and Weisstein, 1978) (see Figure 2.2). This suggests that the computation of one contour representation can depend on the computation of another. In some cases, this is because the sensory mechanisms for detecting cues to contours—like points of contrast in the retinal image—are interdependent. Plausibly, representations of continuous contours, closed contours, and three-dimensional boundaries have as constituents representations of contour

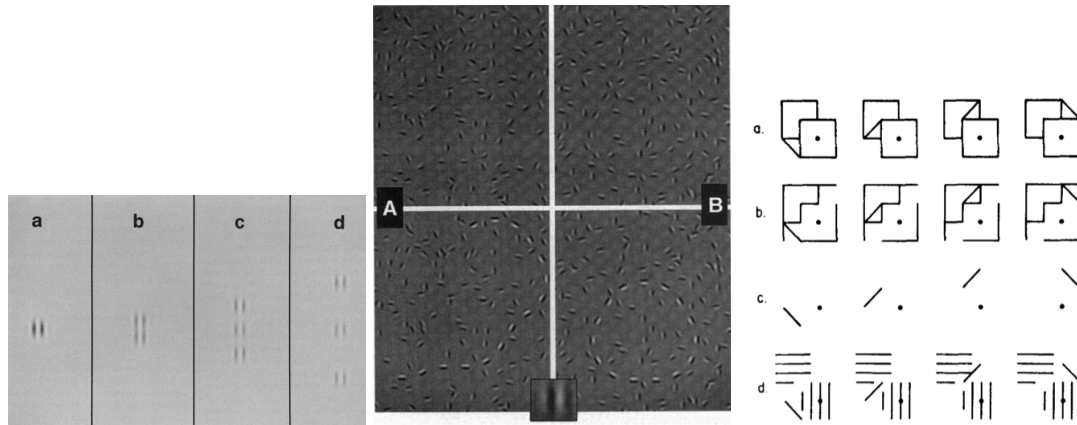


Figure 2.2: Left (from Polat and Sagi, 1993, p. 994): Discrimination thresholds for a contour segment are lowered (that is, discrimination is better) when it is neighbored by collinear contour segments that are higher contrast and that are not too close to the target contour segment (c and d). Center (from Kovács and Julesz, 1993, p. 7495): The closed contour in the upper right (highlighted in the lower right) is more discriminable than the contour in the upper left (highlighted in the lower left), although both contours are composed of segments with the same orientations, spacing, and degree of collinearity. Moreover, an individual element of the closed contour is more discriminable than an identical individual element of the open contour. Right (from Williams and Weisstein, 1978, p. 86): Subjects are better at identifying which of the line segments shown in row *c* occurred when the line segment was presented in the context of a three-dimensional configuration (row *a*) than when it was presented in any other context and even in isolation.

segments. But rather than independently producing representations of individual contour segments and then constructing a complex representation out of them, the production of a representation of one contour segment seems to depend in part on the production of representations of other nearby contour segments.

If we are careful about what the principle of semantic compositionality is, it should become clear that there is no real inconsistency here with compositionality. Compositionality implies that constituents are *semantically* independent, but not that they are *causally* independent. The principle of compositionality only entails that what a representation denotes is a function of nothing more than what their constituents denote and how those constituents are

related. What we have seen is that whether a representation of a contour segment is *produced* depends on what representations of neighboring contour segments are produced. This is compatible with the compositional account on which the representation of the elongated chain of contour segments, for example, is composed from constituent representations of contour segments, and that those constituent representations are semantically independent of each other. The processing of different representations may be interdependent so long as their semantics is not. So at least one feature of the building-blocks analogy is not essential to compositionality: the constituents of a representation may have causal influences on one another. Semantic compositionality bears on the abstract semantic and syntactic relationships among expressions. Given those abstract relationships, there can be quite a lot of variation in how complex expressions are causally processed.

The point that the causal order of representations need not parallel their semantic order relates to the point, in Chapter 1, that structural hypotheses can generalize across different processing accounts. Since representational structure generalizes over many different types of processing systems, and since compositionality concerns how structure relates to content, it is natural that compositionality should not be strictly committed to any particular conception of how representations are causally related. In particular, that a representation has as constituents, and compositionally depends on, some other representations does not mean that the latter representations are causally independent of each other or even that they precede the whole representation of which they are parts.

Likewise, in linguistics there is little expectation that syntactic and semantic relationships

will reflect the order of psycholinguistic processing.⁴ Interpreting structurally ambiguous sentences in real time, for example, is often a context-sensitive and feedback-dependent process. In parsing the sentence, “The horse raced past the barn fell,” one typically first parses “The horse raced past the barn” as a sentence; as soon as one encounters “fell,” one must backtrack and adopt a new parse on which “The horse raced past the barn” is a noun phrase (meaning *the horse that was raced past the barn*). Further, linguistic parsing proceeds linearly as each word is registered. But linear word order cannot encode the hierarchical structures that syntacticians attribute to sentences. The wh-expression “Who” in “Who did you see Ingrid say goodbye to?” is heard first, but standard generative models hold that the base position of “who” is at the end of the sentence.⁵

It is unlikely that a compositional semantic theory of perception will, in any straightforward way, parallel a mechanistic story of how percepts are generated. Yet the particular way in which the representations are compositional has important consequences for why those representations are processed the way they are, and even for why they are processed interdependently. Consider our representations of elongated contours. It is more likely that one will come across continuous, elongated contours of certain shapes than disconnected tiny contours (Geisler et al., 2001). So an individual representation as of a single tiny contour is less likely to be accurate than a complex representation as of an elongated contour. Indeed, elongated contours have more functional significance to us than isolated tiny bits of contour, since they

⁴“When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed, in some practical or efficient way, to construct such a derivation . . . this generative grammar does not, in itself, prescribe the character or functioning of a perceptual model or a model of speech production” (Chomsky, 1965, p. 9; see also Berwick and Weinberg, 1983).

⁵Thanks to Gabe Dupre for helpful discussion.

attach to or bound the sorts of medium-sized goods with which we constantly have to deal. On most psychological accounts, our perceptual representations as of some elongated contours are semantically composed from our representations of component contour segments. Since complex representations of elongated contours are more likely to be accurate and are more functionally significant than any single representation of a tiny contour segment, the visual system should be biased against producing isolated representations of contour segments and biased toward producing sets of representations of contour segments that can be combined into a complex representation of an elongated or closed contour. Such a bias would require interdependent processing of individual representations of contour segments. The compositional semantics of contour representations, together with statistical patterns in the environment, helps to explain why computations of those representations should be interdependent.

Not only is perceptual processing often interdependent, but the computation of representations of complex chunks of the world can often precede or coincide with the computation of perceptual representations of the parts of those chunks. Consider the research stemming from David Navon (1977)'s study of *global precedence*. Navon questioned whether our perceptual representations of complex stimuli are built up and accessed piece by piece from our representations of the simpler elements of those stimuli. Instead, he suggested that we often begin by representing the complex patterns of the world and then parsing them. Navon used the terms "global" and "local" to refer, respectively, to superordinate and subordinate levels of structure. He and his successors found a great deal of evidence that under certain conditions we represent complex stimuli as a whole prior to or simultaneously with representing the individual parts of those stimuli (Kimchi, 2015).

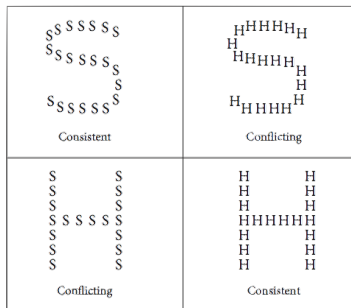


Figure 2.3: “Hierarchical stimuli.” (From Navon, 1977).

Many of the experiments conducted by Navon and his successors use “hierarchical” or “compound” displays consisting of large patterns arranged from smaller elements, such as a group of letters arranged in the form of another letter (Figure 2.3, for example) or a group of shapes of one sort arranged in a shape of another sort. Several types of result are common. First, subjects tend to be faster and more accurate at identifying global patterns than local elements. Second, in cases where the global pattern “conflicts” with the local pattern (for example, an S made out of Hs rather than out of smaller Ss), there are costs to the discrimination of local elements but not of global elements. In other words, global patterns tend to interfere with our discrimination of local elements but not conversely. Third, by presenting a stimulus and then masking it at varying times so as to cut off processing, researchers have found that global patterns support priming before local patterns are able to. These studies all suggest that representations of global patterns can precede and causally influence representations of the local elements in those patterns.

I want to make two points here. The first is that we can give compositional accounts that are consistent with global precedence. The second is that even if the causal story does not mirror the semantic story, the semantic story offers crucial insights into perception. Indeed,

we cannot know what is actually going on in cases of global precedence without a better understanding both of the structures of the representations involved and of their semantic composition.

One possibility that is consistent with compositionality is that, in fact, the perceptual representation of the complex pattern is not structurally related to the perceptual representations of the component elements. Perhaps the temporally later representations of local elements in the hierarchical stimulus (the small “H”s) are not in fact constituents of the prior representation of the global pattern (the large “S”). This possibility is reflected in spatial frequency accounts of global precedence. These accounts hold that representations of patterns at larger spatial scales are computationally prior to representations of patterns at smaller spatial scales. The latter are not constituents of the former.⁶

Another possibility is that the complex representation of the global pattern is composed from constituent representations of elements of that pattern. Global precedence effects may occur because the visual system tends to form the complex representation all together as a whole chunk, so to speak, before attention, memory, recognition, and other such processes can operate on the constituents independently. On this account, the global perceptual representations are formed simultaneously, but not before, their constituents. If the whole representation of the global pattern is formed, then that representation must contain the constituent represen-

⁶Navon supposed that our representations of complex patterns are structured from constituent representations of the components of those patterns (see also Palmer, 1977). On this assumption, global precedence involves processing first the whole representation of a pattern and then the constituents of that representation, which each represent elements of the pattern. But this assumption may not always hold. Navon and other researchers on this topic commonly switch back and forth between describing global precedence as showing that complex representations are processed before their constituents and as showing that representations of complex patterns are processed before representations of sub-patterns.

tations of the local elements. So why is it that the contents of these constituent representations are not available as soon as the content of the global representation? It may be that in order for perceptual processes to operate on the contents of those constituents the processes must first operate on the whole complex (Navon, 1983; Ankrum and Palmer, 1991). On this account, complex perceptual states and their constituents are produced in parallel, but are operated on in order from complex to simple. The situation may be similar to the case of the interdependent processing of contour representations. Because we are typically more concerned with whole objects and configurations than we are with their component elements, the visual system may be biased toward generating complex representations before any constituents can be operated on independently. This account demonstrates another aspect of the building-blocks analogy that is not essential to compositionality: the complex representation need not be built up, piece by piece, from its constituents. The complex representation may be formed all together, as a whole. It may be only later that the constituent parts of the representation are dealt with independently by further psychological processes.

A third possibility is that the perceptual system can generate *abstract schemas* or *frames* with slots or variables in place of constituents. These schemas might determine which modes of combination apply to categories of constituents—for example, which grouping principles govern their combination—without determining the specific identities of the constituents (cf. Love et al., 1999). The schema might carry some abstract content about the type of pattern being represented—for example, that it is a large “S”-shape—and might be progressively fleshed out over time. In this way, the computation of the complex perceptual state might really precede its constituents. This hypothesis makes substantial claims about the structural

and semantic apparatus of perception. The hypothesis assumes that contentful perceptual representations can occur and be operated upon with variables in place of constituents and that the perceptual system can substitute constituents for those variables.

These three compositional hypotheses yield different causal interpretations of global precedence, even though on none of these hypotheses does the causal generation of representations mirror the semantic order of composition. It is reasonable to think that global precedence is not a unified kind of phenomenon, and that different kinds of structures and compositional principles are operative in different cases of global precedence (cf. Pomerantz, 1983). Different compositional accounts may characterize different forms of global precedence. Two aspects of these compositional accounts are especially important for understanding the nature of global precedence. First, these accounts give different characterizations of the structures of the representations involved. Insofar as global precedence involves structure-sensitive processes, different kinds of structural relations require different kinds of processing. Second, the compositional account of the contents of the representations is needed in order to explain global precedence. Global precedence is in the first place defined semantically in terms of the causal relationship between representations of complex patterns and representations of the parts of those patterns. So for the different accounts of representational structure to illuminate global precedence, they must be tied to accounts of how the contents of the representations compose.

I conclude that the semantic compositionality of perception is consistent with processing that does not parallel the compositional order of representations. Perhaps more importantly, compositional semantic theories can be indispensable to our understanding of how perceptual

states are processed, even if the order of processing does not parallel the order of semantic composition.

2.3 Emergent Content

Consider our representation of the tabletop as yellow and rectangular. This representation is structured from—is feature-bound out of—representations as of yellowness and as of rectangularity. The denotation of the whole perceptual state is straightforwardly derived from the denotations of its parts. In most cases, however, the relationship between what complex perceptual states represent and what their parts represent is less direct.

For one, complex perceptual representations will very often represent different determinate features or relations than those represented by their constituents. For example, suppose our representation of the shape of a complex object is structured from constituent representations of the shapes of parts of that object, as Marr and Nishihara (1978) and Biederman (1987) proposed. Suppose also, with these authors, that each of these representations represents shape partly by attributing an axis of symmetry. There are cases where the axis that we attribute to the whole shape does not correspond to the axis that we attribute to any of the components (for example Figure 2.4a).

In a different sort of case, the complex representation represents types of determinable features that are not represented by its constituents. Plausibly the representation of a closed boundary (the square in Figure 2.4b, for example) is structured from constituent representations of contour segments (Richards et al., 1986; Geisler and Super, 2000; Kellman et al.,

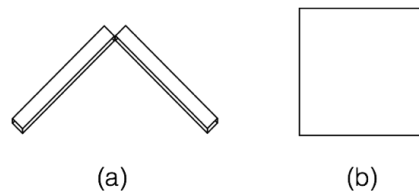


Figure 2.4: (a) A complex perceptual representation of the whole figure attributes an axis to that figure which is not identical to that attributed by any of its constituents. (b) A complex perceptual representation of the square may attribute closure and area—features that its constituents could not attribute to the parts of the square.

2013). The complex representation, on many accounts, represents as of a *closed* contour, with a certain *area* and *aspect ratio*. But none of the constituent representations as of one-dimensional contours could attribute two-dimensional properties like closure, area, or aspect ratio to those contours. As Stephen Palmer wrote,

Whole figures do seem to have natural parts, yet there are properties of the whole which its parts do not share. For example, it is eminently reasonable to believe that the perceptual representation of a square includes the representation of lines as sub-parts. But it is also important to realize that the square has attributes of closedness and area that are not attributes of the component lines. (Palmer, 1977, p. 442)

Closure seems to first appear in the content of the complex representation.

So we have two sorts of cases. First, a complex perceptual representation might represent the same kind of determinable property as its constituents (for example, size, or axis of symmetry), but represent a different determinate than any of its constituents do. Second, a complex perceptual representation might represent a type of determinable property that none

of its constituents do (for example, the closure of a contour). These cases may seem to challenge compositionality, since the whole representation has semantic properties that are not possessed by their constituents. However, careful reflection shows that such emergent semantic properties are explicable in a compositional framework (see Pelletier, 2012). To compositionally explain why a complex representation has the content it has, one can appeal to both the contents of that representation's constituents and to the way in which those constituents are combined. The mode of combination itself may be associated with the representation of emergent properties.

Cases of emergent content violate very simple compositional accounts on which the complex representation must simply *inherit* the content of its constituents. For example, in representing a global axis of symmetry or the closure of a surface's bounding contour, the compositional rule cannot simply be to take the intersection of the contents of the constituents. In fact, the compositional rules in these cases typically cannot be primitive logical or set-theoretic operations (conjunction/intersection, disjunction/union, negation/complementation) on the contents of the constituents. A conjunction of representations of the sides of a square, for example, would not represent a whole closed figure, but would rather represent such-and-such a line *and* such-and-such a line *and* such-and-such a line *and* such-and-such a line. As Max Wertheimer (1938a) emphasized, one cannot analyze perceptual representations of global properties such as closure in terms of mere "and-summations" of component percepts.

Cases of emergent content do not in any obvious way violate compositionality. They do demand complex compositional explanations. The compositional principles for emergent contents must encode general facts about how global symmetries can be derived from component

symmetries, or about how properties like closure, area, and aspect ratio can be derived as a function of the relations between contour segments (see, for example Elder and Zucker, 1994). These compositional principles must encode substantive geometrical rules, and it can be hard work to determine what those rules are. However, specifying those rules helps to illuminate how representations of local bits of the world—parts of objects, fragments of contour, and so on—can compose into representations of whole discrete objects and surfaces with global properties.

The standard refrain of Gestalt psychology is that the whole perceptual representation is other than the sum of its parts. Interpreted broadly, this means that the whole perceptual representation cannot be analyzed in terms of its parts and their mode of combination. If interpreted in this way, the claim is unfounded. Still, a narrower interpretation of the Gestalt dictum is warranted: the whole is not a *sum*—or any other simple function—of the parts and their mode of combination.

2.4 Context Effects

A context effect arises when the way we represent one element in a scene depends on our representations of other distinct elements in the scene.⁷ Where there are context effects, it may seem that the content that a perceptual state carries about the world depends on how that state is embedded in a more complex perceptual representation. This would violate compositionality,

⁷By “context” I exclusively mean the representations with which a representation is combined (for example, the “sentential context” in the case of language). I do not mean the part of the world in which the representation is had (for example, the “utterance context” in the case of language) or the part of the world in which the represented object is situated.

which implies that the content of a perceptual state can only depend on the contents of its constituent states and how they combine.

The problem may be highlighted if one thinks of perceptual representations on the model of pictures. Katalin Balog has pointed out that

It is not at all clear. . . that pictures are compositional in the way sentences are. The part of The Last Supper that represents Judas' eyes does so only in the context of the whole picture. A matching bit of canvas in a different painting might represent something quite different (say the marking on a snake). (Balog, 2009, p. 315)

What could that isolated pattern of paint depict, independently of its context? To make the point even more vivid, consider a single pixel in a digital picture, or a spot on a Seurat painting. What could *they* represent, abstracted from their contexts?

We can visually represent features of the world that project patterns of light as small as, or smaller than, the receptive field of a photoreceptor (Westheimer, 1981). How could these perceptual states possibly carry content about the world independently of their representational contexts? I will show how we can, in fact, give compositional accounts of context effects in perception. I will focus on three cases: the perception of lightness and color, the perception of three-dimensional edges, and the interaction between our perceptions of orientation and shape. I choose these cases for three reasons: they are analogous to context effects we find in pictures, it is easy to demonstrate the perceptual context effects, and they provide opportunities for illustrating different kinds of structural and semantic accounts. There is substantial work to be done in precisifying and empirically testing these accounts.

Before we turn to perception, however, I will set up the problems for treating context effects by looking at analogous problems in natural language.

2.4.1 Context Effects in Language

Higginbotham (1986) presented context effects in language as counter-examples to compositionality. Since the problem posed by context effects in perception is analogous to the problem posed by context effects in language, we can get a sense of how to proceed by looking at how Higginbotham's cases work and some possible responses.

Consider the following sentences (from Pelletier, 1994, p. 602):

(1) John will eat steak unless he eats lobster.

(2) No person will eat steak unless he eats lobster.

(1) is true if John eats steak but not lobster. Here, "will eat steak unless he eats lobster" seems to mean something like *x eats steak OR x eats lobster*. (2) is true only if everyone who eats steak also eats lobster. Here, "will eat steak unless he eats lobster" seems to mean something like *x eats steak and x does not eat lobster*. Higginbotham describes this as a case where the noun phrase, "John" or "No person," partially determines the meaning of the verb phrase, "will eat steak unless he eats lobster." Yet that noun phrase is not a constituent of the verb phrase. If a language is semantically compositional, then the meaning of an expression can only depend on what its parts mean and how they combine; the sentential context of an expression cannot play any role in determining the meaning of that expression. So, if Higginbotham's description is correct and the external noun phrase influences the meaning of the verb phrase, then English

is not compositional.

There are several strategies for redescribing such context effects in a way that aligns with compositionality (see Pelletier, 1994; Janssen, 2011). In the case of (1) and (2), one might propose that there are really two homonyms, “unless₁” and “unless₂,” each with different meanings, and that the noun phrase selects for the verb phrase containing one or the other term. Or one might hold that the string “eat steak unless he eats lobster” has different syntactic structures—is characterized by different modes of combination—in (1) and in (2). These different structures may be associated with different semantic interpretations. For example, “unless” may take different scopes in relation to the noun phrases in (1) and (2).

Finally, one might hold that “unless” has a more abstract meaning than we initially supposed. There are several ways in which the term might have a more abstract meaning. It may have a disjunctive or piecewise meaning. Or the term’s meaning may be a function from some parameter whose value is specified by the noun phrase to a particular connective. Or the term’s meaning may simply be vague or indeterminate. *some connective or other amongst this set...*, for example (Pelletier, 1994). Then the choice of noun phrase does not determine the meaning of the verb phrase, but rather combines with the verb phrase’s meaning in determining the meaning of the whole sentence. So the verb phrase might have the meaning *x eats steak [some conjunction] x does not eat lobster*. By virtue of being combined with “No person” (the details do not concern us here), the derived meaning for the whole sentence is that *no person eats steak and does not eat lobster*.

In general, we can always preserve the commitment to compositionality by either propos-

ing a distinction between primitive representations with different contents, or by distinguishing different modes of combination along with an attendant difference in semantic rules of composition, or by positing some more abstract or indeterminate content that combines with additional parameters supplied by the content of some other representation, so that the resulting whole has a more determinate content (see Janssen, 2011, p. 516–7).

These theoretical adjustments are beholden to the linguistic data. For example, the different compositional treatments of “unless” are non-trivial, empirical hypotheses. The claim that the occurrences of “unless” in (1) and (2) are homonyms predicts that “unless” should behave like other lexically ambiguous terms such as “bank.” The claim that the verb phrases in (1) and (2) have different syntactic structures should predict similar structural ambiguities in other sentences. And a theory on which “unless” has a more abstract meaning should generalize to all occurrences of “unless.”

Compositionality prescribes a number of options for handling context effects. But which of these is correct, if any, is an empirical matter. Likewise, I will argue that there are a number of ways to handle perceptual context effects in a compositional way. These different accounts will have different testable consequences.

2.4.2 Lightness and Color

Context effects are common in lightness and color perception, in which the perceived lightness and color of a patch depends on our perception of surrounding patches. Consider the simultaneous contrast effect, illustrated on the left of Figure 2.5. The nature of the light reflected by



Figure 2.5: Left: The simultaneous contrast effect. The central patches reflect the same light, but look to have different lightnesses as a result of their different surrounds. Right (from Adelson, 2000, p. 344): the indicated patches give off the same light but look to have different lightnesses as a result of how we perceive them to be part of the spatial organization of the surface.

each central square is the same. Yet the central square on the left will tend to appear lighter than that on the right. The perceived lightness of a surface also depends on how the perception of that surface is grouped with other surfaces. For example, two patches on the corrugated surface on the right in Figure 2.5 give off the same light. Yet whether we attribute the same or different lightnesses to these patches depends on whether or not the perceptual representations of those patches combine into a perceptual representation of a common plane (Adelson, 2000; Gilchrist, 2015).

Elizabeth Schier argues on the basis of context effects like these that “It is not possible to consider the representational content of my red experience in isolation from the other representing vehicles because it only has its representational content in virtue of its relations to the other [non-constituent] vehicles” (Schier, 2007, p. 20). She writes, “an individual vehicle only has meaning because of its place in the overall structure” (Schier, 2007, p. 18). Taken at face value, these statements present a non-compositional description of color and lightness perception, since the content of a perceptual representation is taken to depend not on its *internal structure* but rather on how that representation is combined with other non-constituent

representations. Compositionality, by contrast, requires that the content of a representation depend only on the contents of its constituents and how they are put together.

It is worth pointing out, in the first place, that Schier's description of these context effects is puzzling. She suggests that the *very same* type of color representation could have different content in different contexts. But what are the identity conditions of this type of representation? One cannot say that the representations are type-identical in virtue of their phenomenal character, because the perceptual states do not have identical phenomenal characters. What it is like to view the central square in a dark surround is different from what it is like to view that same square in a lighter surround. And these perceptual states cannot be type-identical in virtue of their content, since on Schier's own description they do not have any content considered in isolation. Further, Schier does not submit any neurophysiological or functional way of type-identifying the perceptual states. The description according which these perceptual states have different contents but are identical in kind is difficult to interpret.

Schier's primary aim was to characterize color representations in a way that accounted for the fundamental interdependencies in how one perceives the colors of different surfaces. I believe one can give a better characterization that nevertheless preserves the basic spirit of her account within a compositional framework. On one compositional account, the perceptual representations of the central patches in Figure 2.5a are type-identical according to their content. One could posit that the representations have the same content, but that this content is more abstract than a specific color or lightness. For example, one common explanation of the simultaneous contrast effect is that the surround offers a cue to how the central patch is illuminated. A lighter surround suggests a lighter illumination. The same amount of light is registered for

the central squares in each surround. That amount of light must be reflected either by a well-lit, darker surface or by a darkly-lit, brighter surface. So, if the surround presents a cue that the illumination is brighter, then we will perceive the central patch as correspondingly darker.

One could propose that the content of the representation of the central patch is a function from an illumination value to a lightness value, such that the illumination and lightness values combine to produce the light registered at the eye.⁸ When a representation of the central patch is combined with a representation of a darker illumination, we get a complex representation of a scene containing a central patch of a determinate lightness. When the representation of the central square is combined with a representation of the lighter illumination, we get a complex representation of a scene containing a central patch of a determinate, darker lightness.

Whereas Schier's account offers little understanding of how the representations of the central squares are identical in kind, the compositional alternative offers a bit more insight, ascribing the same contents to the representations. The assumption, common to both approaches, that the representations, *modulo* context, are identical in kind is an empirical matter. One would want evidence that the visual system treats those local representations of the central patch as identical in kind. Of course, there will be *sensory* states in common when we look at the central patches. By design, the central square in each figure should produce the same spectral intensities of light at the eye. What we want to know is whether there are *representational* states in common (see Burge, 2010a). In the absence of reasons for thinking that this is

⁸Where i is an illumination value and l is a given irradiance registered at the eye, on this account the content of a perceptual representation of an individual surface patch is $\lambda i \cdot \frac{l}{i}$. Suppose the visual system derives a representation of the illumination i from a representation of the surrounding square. The result of combining a representation of a surface patch with a representation of illumination i is a state that represents $(\lambda i \cdot \frac{l}{i})i$.

so, the claim that the representations are identical in kind is unmotivated. If there is evidence of type-identical representations of the central patch, we still must evaluate whether specific claims about the content of those representations yields good predictions about perceivers' performance on psychophysical tasks.⁹

An alternative route to compositionality is to posit that the representations of the central squares are simply different in kind and that they have different contents. Perhaps one perceptual state represents a determinate lightness l_1 and the other represents a determinate lightness l_2 . The context effect, on this account, is an artifact of structural constraints that bias certain combinations of lightness representations over others. Which representation of the center is tokened—whether it is the state that represents l_1 or a distinct kind of state that represents l_2 —depends on which representation of the illumination is being produced. But once tokened, the content of a perceptual state can be specified independently of the content of the other states that are generated alongside it.¹⁰

A more subtle and semantically interesting account may be drawn from *anchoring* theories of lightness and color perception (for example, Gilchrist et al., 1999). On one interpretation of anchoring theories, the visual system makes representations as of absolute lightness or color on the basis of representations as of relative lightness or color together with a representation of a particular surface as having an absolute lightness or color. The surface to which we primitively attribute an absolute lightness or color is called an *anchor*. Suppose that

⁹In "The Perspectival Character of Perception," I argue against attributing representational contents of the sort described in the previous paragraph and in Ftn. 8.

¹⁰Cf. Byrne and Hilbert (2003), who argue for an *absolutist* metaphysics of color, according to which color is a non-relational property. Byrne and Hilbert explain context effects as cases where context is among the conditions necessary for perceiving a given color, but where that context is not part of the nature of color.

one's representation of the central patch represents that patch as having a *lightness relative to its anchor* (the anchor, in this case, being the surrounding square). For example, one will represent the central patch in the darker surround as having twice the lightness of its anchor, while one will represent the central patch in the lighter surround as having 85% the lightness of its anchor. These relative lightness values will be computed on the basis of the differences in registered light intensities on the retina, along with heuristics for determining which surface will be represented as the anchor (see, for example Land and McCann, 1971). On this account, the representations of the central patches have different contents, though they do not attribute absolute lightnesses. Given sufficient cues, we can attribute an absolute lightness to the anchoring surface—that it is white, that it is black, or that it is midway between white and black, for example. If one can combine a representation of the central patch's relative lightnesses with a representation of the anchor's absolute lightness, the resulting complex will attribute absolute lightness values to both the surround and the central patch.

On this sort of account, individual lightness and color representations do have their contents independently of their contexts, but this content is more abstract than reference to some absolute lightness or color. The content of such a representation may contain parameters—the lightness of the anchor, for example—whose values are supplied upon combining with other representations. Only complex representations of both center and surround will represent the center as having a particular lightness or color.¹¹

¹¹This sort of semantics is consonant with *relationalist* views of the metaphysics of color (cf. Cohen, 2004), according to which colors are relational properties that involve, among other things, surrounding surfaces. However, the semantic and metaphysical accounts are not tied to each other. One could hold that color is a non-relational property, but that only complex representations of multiple surfaces can attribute those non-relational properties to particular surfaces.

An analogous semantics is quite natural for theories of *cue combination*. For example, sensory cues to depth differ on what types of information they supply. Some cues allow us to represent relative depth—that one surface is three times farther away than another, for instance. Landy et al. (1995) describe a model of cue combination on which representations of relative depth combine with representations of absolute distance to a surface in order to yield a representation of absolute depth. A perceptual state will represent a surface *b* as 30 units away from the perceiver if that state is a combination of a representation of surface *b* as three times farther away from surface *a* with a representation of surface *a* as 10 units of distance from the perceiver. In general, Landy et al.’s model supposes that different sensory cues give rise to depth representations that contain certain parameters the values of which can be supplied upon combination with other representations.

One consideration in favor of some relational semantics, on which we represent relative lightness or color, is its ability to account for both *relational* lightness and color constancy and *absolute* lightness and color constancy (see Foster, 2011). *Relational lightness and color constancy* refers to our ability to perceive a surface’s relative lightness or color under different illuminations, even if we are unable to represent its absolute lightness or color. We may not be able to represent absolute lightness or color when, for example, there are not adequate cues for the absolute lightness or color of the anchoring representation. The relational semantics gives a natural account of the contents of these representations of relative lightness and color, and of how these representations can contribute to representations of absolute lightness or color when combined with available representations of the absolute lightness or color of the anchoring surface.

Context effects in lightness and color perception can be treated compositionally and with less mystery than non-compositional treatments. I have sketched a few compositional accounts, without intending to exhaust the alternatives or decisively label one as offering the best explanation of context effects in lightness and color perception. These accounts have different empirical and explanatory upshots. There is a great deal of work, both philosophical and psychological, left to be done in further articulating and testing these accounts.

2.4.3 Three-Dimensional Edges

Let's turn to another case. Context effects are common in the perception of edges. Different types of edges, or contour segments, in the world hold perceptual significance. Let's focus on three types: edges that arise at the point where a surface occludes itself; convex or outside edges that arise when two surfaces meet in your direction; and concave or inside contours that arise when two surfaces meet away from you. These different edge types correspond to different ways that surfaces can be configured in three-dimensional space with respect to a viewer. Crucially, how we perceive a given edge—whether as occluding, convex, or concave—depends on how we perceive other connected edges.

For example, suppose you are monocularly viewing three-dimensional scenes that project the images in Figure 2.6. The edges a and i , on their own, project the same type of image. Yet we tend to perceive a as a convex edge while we perceive i as an occluding edge. The difference in our perception of a and i depends on our perception of n . When the visual system registers two lines in the retinal image that are arranged in a “T” pattern, it typically infers that the contour that projected the top of the “T” is an occluding contour that belongs to

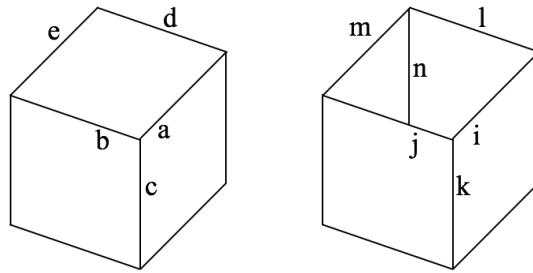


Figure 2.6: We tend to perceive *a* as a convex intersection between surfaces and *i* as a self-occluding contour.

a closer surface while the contour that forms the stem of the “T” belongs to a farther, partially occluded surface (Shimojo et al., 1989; Enns and Rensink, 1991; Kellman and Shipley, 1991; Rubin, 2001). Since *n* and *j* project such a T-junction to the eye, we are likely to perceive *j* as an occluding contour. Notice that the point at which *i*, *j*, and *k* meet forms a “Y” pattern. As a general rule, if we perceive the upper left edge in a “Y” pattern as occluding the inside of a hollow object, then we will also perceive the upper right edge as occluding the inside of that object. So, in virtue of our perceiving *j* as occluding and *i* as forming a Y-junction with *j* and *k*, we perceive *i* as an occluding edge. How can a compositional semantics account for these context effects?

Here is a non-compositional way to characterize this case. We form representations of each individual contour. What content a given edge representation carries depends on how that representation interacts with representations of other edges in the figure. Whether a representation of an edge represents that contour as occluding, convex, or concave depends on how it is combined with representations of other edges. The content carried by a contour representation is a function of its relation to other non-constituent edge representations.

But there are also compositional ways to characterize the case. To make this perspicuous, let's adopt some notation. I don't intend to offer a strictly regimented meta-language for describing perceptual representations and their semantics. Yet it will be convenient to have at hand some rough and ready notational conventions. I will use italicized, lower-case letters to denote types of particulars in the world. Bolded letters denote perceptual states that represent as of those types. So, on the intended interpretation, \mathbf{x} is to be read, "A perceptual representation as of x ," where x may be a type of edge, color, shape, or what have you. Different subscripts will mark different token representations and token objects. Different superscripts or primes will mark states that represent as of the same item but have different structures. So \mathbf{a} refers to a representation as of a , and \mathbf{a}' refers to a structurally different representation as of a . I will use non-bolded letters to denote structural modes of combination, allowing us to specify the structure of some representation \mathbf{x} as, for instance, $F(\mathbf{y}, \mathbf{z})$. And I use $\llbracket \mathbf{x} \rrbracket$ to refer to the content of a representation \mathbf{x} . On the intended interpretation, \mathbf{x} denotes a percept that represents x in some way, and $\llbracket \mathbf{x} \rrbracket$ denotes the way \mathbf{x} represents x . For example, for Figure 2.6, $\llbracket \mathbf{a} \rrbracket = \text{convex}(a)$.¹² One way to specify a compositional semantics is to give rules of the form: $\llbracket F(\mathbf{x}, \mathbf{y}, \mathbf{z}) \rrbracket = f(\llbracket \mathbf{x} \rrbracket, \llbracket \mathbf{y} \rrbracket, \llbracket \mathbf{z} \rrbracket)$.¹³

Now we are in a position to look at compositional treatments of context effects in our perception of three-dimensional edges. One option, as we saw in the case of lightness, is to take the representations of a and i to have a more abstract content. On such an account, the specific content that an edge is of one type or another really emerges through semantic composition

¹²I will largely bracket discussion of the singular referential elements in perception, as my main concern is on how the attributive elements of perception are semantically composed.

¹³A semantics that contains only compositional rules of this form is *locally* compositional (see Footnote 3).

as the representation of that edge combines with representations of other edges. For example, suppose that the representation of a carries the content that the edge is a *surface event*—that is, an abrupt change in the surface. Or, along similar lines, suppose that $\llbracket \mathbf{a} \rrbracket = \llbracket \mathbf{i} \rrbracket = \{+, -, >\}$, where “+” abbreviates *convex*, “-” abbreviates *concave*, and “>” abbreviates *occluding*. We might posit that there are different ways of combining edge representations, corresponding to different ways that the edges can intersect. The different modes of combination correspond to compositional rules that further narrow down the nature of the edges denoted by the constituent perceptual states. Typically, as representations of edges combine into representations of intersections, and as representations of intersections combine with each other, one will get more determinate content about the nature of the different edges.¹⁴

For example, here is a simplified fragment of a semantics for the context effect. Suppose i, j, k are constituents of a representation $Y(i, j, k)$. We determine the content of $Y(x, y, z)$ using something like the following rule:

$$\begin{aligned} \llbracket Y(\mathbf{x}, \mathbf{y}, \mathbf{z}) \rrbracket = \{ & s \in \llbracket \mathbf{x} \rrbracket \times \llbracket \mathbf{y} \rrbracket \times \llbracket \mathbf{z} \rrbracket : s = (>, +, >) \vee s = (>, -, >) \vee s = (+, +, +) \\ & \vee s = (-, -, -) \} \end{aligned}$$

Let’s take for granted that $\llbracket \mathbf{i} \rrbracket = \{+, -, >\}$. And suppose that since j corresponds to the top-bar of a T-junction in the retinal image, the visual system has generated a representation j

¹⁴Semantic composition, on this account, amounts to what is known as “constraint propagation” (see Waltz, 1975; Reiter and Mackworth, 1989). One might think of each primitive representation as a variable that takes a limited set of possible values. Combinations of representations place constraints on which values can be assigned simultaneously to each representation.

such that $\llbracket j \rrbracket = \{>\}$.¹⁵ Then

$$\begin{aligned} \llbracket Y(i, j, k) \rrbracket &= \{s \in \llbracket i \rrbracket \times \llbracket j \rrbracket \times \llbracket k \rrbracket : s = (>, +, >) \vee s = (>, -, >) \vee s = (+, +, +) \\ &\quad \vee s = (-, -, -)\} \\ &= \{(>, +, >), (>, -, >)\} \end{aligned}$$

(Other rules must be formulated to explain why k is represented as convex.)

So the representation of i by itself has underspecified or abstract content about the type of three-dimensional edge that i is. The combined representation of i , j , and k represents i specifically as occluding, given that j is represented as occluding and the rule for interpreting Y -form representations contains something like the above principle. This account is really just an illustration, of course. There may be other ways to implement the general idea that the perception of an edge's nature emerges through semantic composition.¹⁶

¹⁵One might be tempted to posit a mode of combination by which the representations j and n are combined into $T(j, n)$, where $T(j, n) = \{s \in \llbracket j \rrbracket \times \llbracket n \rrbracket : s = (>, -) \wedge s = (>, +)\}$. However, it seems to me more plausible that our representations of n and j do not combine with each other into a single representational unit. A T-junction offers a sensory cue that, for example, j and n belong to separate surfaces, with j occluding n . I think it is implausible that the visual system would combine representations of edges that it locates as belonging to different surfaces. I think it is more likely that T-junctions are merely features of sensory stimuli, which have the effects that (i) we primitively represent the edge projecting to the top-bar (j) as occluding, and (ii) we *do not* immediately combine our representations of that edge with our representation of the edge projecting the stem (n).

¹⁶The rich literature on “line labelling” and the computational interpretation of line drawings has tried to capture the principles, like the one given above, by which we perceive edges as occluding, convex, or concave—not to mention whether we perceive them as reflectance edges or illumination edges (see Guzmán, 1968; Clowes, 1971; Huffman, 1971; Waltz, 1972; Sugihara, 1986). The problem, as it is treated in the computer vision literature, is to specify the rules by which a representation of edges can be translated into a representation of symbols designating occlusion, convexity, or concavity. Instead of treating those rules as rules of translation, from one set of representations to another, the above semantic approach treats those rules as giving the compositional principles for deriving the contents of complexes of edge representations.

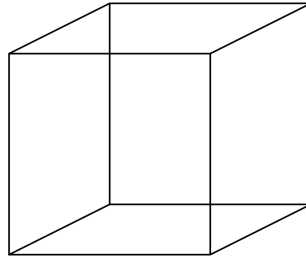


Figure 2.7: Our perception of the Necker cube is bi-stable. We can alternate between seeing it as facing upward to the right or as facing downward to the left.

There are empirical problems for this type of explanation. Consider the “Necker cube” in Figure 2.7. Our perception of the Necker cube is bi-stable. We can alternate between seeing it as facing upward to the right or as facing downward to the left. On the hypothesis that representations of edges have underspecified content, $\{+, -, >\}$, it would seem that all the representations of the edges of the Necker cube should compose into a complex representation that has underspecified content, which I will abbreviate: $\{upward-facing, downward-facing\}$. But this does not explain why our perception of the cube is bi-stable. Instead, it predicts that we should have an abstract, underspecified representation of the orientation of the cube. In other words, we should have a stable perceptual representation of the cube as having an indeterminate orientation, rather than two representations of the cube as having different determinate orientations.

One response might be to posit a rule to the effect that the content carried by every complex perceptual representation must be more specific, or determinate, than the contents of its constituents.¹⁷ In cases where the content of a perceptual representation might be made

¹⁷One could call this the “Principle of Greatest Commitment,” in contrast to Marr’s “Principle of Least Commitment”:

It is frequently the case during the execution of a recognition task that there are a number of

more specific in different ways, the visual system switches back and forth between these more specific contents. However, this rule would not be compositional since the content of the representation of the Necker cube would not be a *function* of its internal structure: the rule relates the representation of the Necker cube to more than one content.

Aside from whether the rule to select a specification of a percept's content is compositional, it is clear that such a rule does not hold generally in perception. For example, when one perceives a surface as occluding another, one perceives the former as nearer to oneself than the latter. The perception of occlusion cues a representation of ordinal depth relations. If one only has occlusion as a cue to depth, the visual system will not make a more specific commitment about how much farther away one surface is than the other. As another counter-example to the proposed rule, consider certain demonstrations of the "phi phenomenon" (also called "object-less motion"). One is presented with a ring of luminous discs. In very quick succession, each disc "flickers off" so that in place of the disc you simply see the background. As the discs flicker off in quick succession, one perceives there to be an occluder sweeping around and covering up each disc. But since the occluder has the same color as the background, one does not perceive the occluder to have any specific shape or size (Steinman et al., 2000). The visual system does not appear to force a more specified representation of the occluder's shape and size. One would have to explain why a commitment to determinacy is a rule in the perception of the Necker cube but not in the perception of depth or in the phi phenomenon.

possible interpretations of a particular datum, but that there is not yet sufficient evidence to decide between them. In such cases, one should never become committed to one of the possibilities prematurely, because of the damage that knowledge associated with that possibility and not with the others can subsequently do. (Marr, 1976, p. 486)

The natural explanation for why our perception of the Necker cube is bi-stable is that the stimulus permits two structurally distinct types of representations—a representation of the cube as *upward-facing* and another representation of the cube as *downward-facing*, but not a representation of the cube that leaves it unspecified whether it is upward-facing or downward-facing. There are at least two ways in which the representations might differ: they might be structured in different ways from the same set of edge representations, or they might be structured in the same way from different primitive edge representations. Let's start with the first option.

Suppose, as before, that primitive edge representations have the content $\{+, -, >\}$. Rather than supposing that there is one mode of combination for representations of edges that intersect in a “Y” pattern, one might suppose that there are multiple such modes of combination Y_1, Y_2, \dots . One may posit constraints on how different types of Y -representations can be combined with other representations as of intersecting edges. For example, suppose $Y_3(i, j, k)$ and $\uparrow_5(l, m, n)$ are the only Y - and \uparrow -structures that can combine. Each mode of combination might correspond to a different semantic rule governing how the complex perceptual state represents the intersecting edges. So the content of $Y_3(i, j, k)$ is given by a semantic rule that entails i 's being an occluding edge. Then one can explain the context effect as a result of the constraint that only a Y_3 representation can combine with a \uparrow_5 representation, and that the semantic rule for a Y_3 representation entails that the upper-right edge is an occluding one. And one explains the bi-stability of our perception of the Necker cube by constraints that just permit a combination of one set of intersection representations or else a combination of another set of intersection representations.

One worry about both this and the previous account is that it is not clear that we should posit perceptual *representations* with the content $\{+, -, >\}$ rather than *sensory states* that merely register patterns in the retinal image that can be projected by edges of different types. Whether we say there are such representations may depend on whether we have reason to posit a perceptual constancy for picking out surface events at particular positions in space. And this in part depends on finding evidence that we perceptually treat edges that project different images—for example, projecting patterns of contrast to different retinal positions and at different retinal orientations—in similar ways, even where those edges are of different types. This can be tested with, among other things, psychophysical “same-different” tasks and priming tasks.

Another alternative is that one’s local representations of a and i in Figure 2.6 are distinct in kind. We may posit that representations as of occluding edges, as of convex edges, and as of concave edges are all primitively different psychological kinds. One only forms representations as of edges of determinate types. If this view is right, the explanation for the context effect could be that a representation as of an occluding upper-left edge cannot combine into a Y -representation with anything but a representation as of an occluding upper-right edge. The T-junction projected onto the retina by n and i biases one toward forming a representation $i^>$. Moreover, $Y(i^>, k^+, j^>)$ and $Y(i^>, k^-, j^>)$ are the only Y -combinations that are structurally possible. One’s perception of the Necker cube is bi-stable, here, because of constraints on how types of primitive edge representations can be combined. Rather than forming a single representation of the Necker cube that has the content $\{upward-facing, downward-facing\}$, there are two structurally possible representations of the Necker cube, one with the content

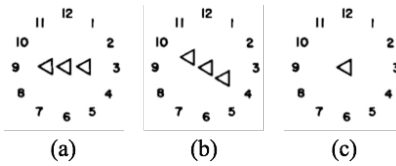


Figure 2.8: Context effects in the perception of orientation: We perceive a triangle as pointing in different directions depending on how it is embedded in a surrounding configuration of triangles. (From Palmer, 1980).

upward-facing and the other with the content *downward-facing*.¹⁸

The representation of three-dimensional edge types can be handled compositionally. Moreover, the different accounts have distinctive empirical consequences. For example, we considered one account on which the context effects are accounted for entirely in the compositional rules. However, that account did not explain the bi-stability of our perception of the Necker cube. We considered two other accounts which differ in whether they treat all primitive representations of edges as identical in kind, or whether edge representations primitively differ in the types of three-dimensional edges they function to represent. There is much room left to precisify and test these different accounts.

2.4.4 Orientation and Shape

Attneave (1968) noted that equilateral triangles can be alternately seen as pointing in one of three different directions. But, he noted, we cannot see them as pointing in all three directions at the same time (see Figure 2.8). The perceived direction of the central triangle depends

¹⁸The first account that we considered interpreted the algorithms posited in the line labeling literature as specifying the rules for semantic composition (see Ftn. 16). The account now being considered assumes that those algorithms specify structural rules on how primitive representations of occluding edges, convex edges, and concave edges can combine.

on several things (Palmer, 1980). First, there is a general tendency to perceive the triangle as pointing in the direction closest to vertical (either upward or downward). One tends to see the triangle in Figure 2.8(c) as pointing upward to 1 o'clock, downward to 5 o'clock, or leftward to 9 o'clock. Second, neighboring figures tend to be perceived as being pointed in parallel directions. Third, if the whole configuration of triangles can be perceived to point in a certain direction, then one will tend to perceive each individual triangle as pointed in a direction parallel with the direction in which the whole configuration points. I will pass over the details of how these tendencies interact.

Let t denote the central triangle in the different configurations in Figure 2.8. The real numbers from 1 to 12 can be used to stand in for the different directions in which a figure might be represented as pointing.¹⁹ These numbers correspond to directions on a clock face parallel to and centered on the surface of the represented figure. Under normal conditions, t could only have the contents 1, 5, or 9 (bracketing whatever content these representations carry about features other than orientation).

The representation of t in (a) has the content 9, while the representation of t in (b) has the content 1.²⁰ If the constituents of our representations of t are the same in kind in (a) and (b) and are combined in the same way to form representations of t in (a) and (b), then the contents of our representations of t in (a) and (b) cannot be derived compositionally. To

¹⁹Or more accurately a quantization of this interval, since our ability to discriminate directions has limited resolution. We also could just as easily have used a quantization of the reals on the closed intervals $[0, 360]$ or $[0, 2\pi]$.

²⁰This should be qualified. One is not guaranteed to perceive t as pointing to 9 o'clock in (a) and 1 o'clock in (b). But it is significantly more likely that one will initially perceive t as pointing to 9 o'clock in (a) and 1 o'clock in (b). For ease of exposition, I will leave this qualification implicit in the main text.

preserve compositionality, one might propose that the representations of t in (a) and (b) have the same content. Or one might propose that the representations are not identical, either with respect to their constituents or with respect to how those constituents are combined. Let's evaluate these options in turn.

Grant, for the moment, that our representations of t in (a) and (b) are type-identical. One can preserve compositionality by positing that t has the same content in both contexts. For example, t may have the content $\{1, 5, 9\}$ or perhaps $\{x : x \text{ is along an axis of symmetry of } t\}$. On this proposal, it is only the representation of the whole array of triangles that represents t as pointing in a particular direction. The complex representation of a whole configuration may carry more specific content about the direction in which a single triangle in that configuration points than is carried by the constituent representation of that triangle by itself. Rules for semantic composition would determine how the specified content of a complex representation is derived from the underspecified content of its constituents. As mentioned above, the visual system prefers (i) to represent isolated figures as directed vertically, (ii) to represent neighboring triangles as pointing in parallel directions, and (iii) to represent triangles as pointing in a direction parallel with the direction in which the total configuration is represented as pointing. These preferences can be modeled by rules of semantic composition.

One might give the following sort of account: each representation of a particular triangle in (b) has the content $\{1, 5, 9\}$, with respect to a clock face centered on that triangle. The representation of the configuration of triangles is a combination of the representations of each triangle. As a rule, the combined representation carries the content that the triangles in the configuration have parallel directions. Since the constituent representations have the content

$\{1, 5, 9\}$, the combined representation will carry the content that the triangles each point at 1 o'clock, or each point at 5 o'clock, or each point at 9 o'clock. And, as a rule, if the configuration can be represented as having a direction, then the elements in the configuration will be represented as pointed in a direction parallel to that of the whole configuration. Since the configuration of triangles in (b) can be represented as pointing at 1 o'clock, each of the triangles in that configuration will be perceived to point at 1 o'clock.

If this is right, then t has the same abstract content in the context of one's representations of the neighboring triangles in (a) and in the context of one's representations of the neighboring triangles in (b). But one's total representation of the array in (b), for instance, will represent t as specifically pointed to 1 o'clock. This story is perfectly compositional. The content of t does not depend on its context. Semantic rules govern how the underspecified content of t contributes to the more specific content of the complex representation of which t is a constituent.

But, as Attneave pointed out, our representation of the triangle in (c) is multi-stable: we perceive it as pointing specifically at 1 o'clock, or specifically at 5 o'clock, or specifically at 9 o'clock. But we do not (at least, not without great cognitive effort) perceive the triangle as pointed in *one or another of these directions*. We switch between perceiving the triangle as having different determinate directions. In addition, Attneave noted that unlike many other cases of multi-stability (e.g. our perception of the Necker cube in Figure 2.7), there would be no logical or geometrical inconsistency in representing the triangle as pointing in all three directions. But we do not (at least without great effort) see the triangle as *pointing at 1 o'clock AND 5 o'clock AND 9 o'clock*. To put it another way, we do not see the triangle as having more

than one base. We neither perceive the triangle as having an indeterminate direction, nor do we perceive it as having more than one direction at once. The hypothesis that our perceptual representation of t has underspecified content does not predict or explain either of these data.

The multi-stability in our perception of the triangle's orientation suggests that our perception of t as pointing in different directions involves our occasioning structurally distinct types of representations of t . What we want then is an account of how the representations of t in (a) and in (b) are distinct. Are they distinct because they are made of different constituents? Or are they distinct because their constituents are combined in different ways? For the sake of space, I will just explore the possibility that the representations contain different constituents.

One natural possibility is that the representations of t contain constituents that function to represent t 's axis of orientation. So, t^a has a constituent $\mathfrak{9}$ while t^b has a constituent $\mathfrak{1}$. Whereas in the previous semantic account we characterized the context effect as a result of semantic rules of composition, we might instead characterize the effect as a result of structural constraints on combination. For example, the visual system is biased toward combinations of representations whose constituents are representations of parallel orientations. These combinations of representations represent the configuration of oriented triangles. The visual system is also biased toward attaching to these configural representations representations of a configural orientation that is parallel to the orientations specified by the constituent representations of figural orientation. Letting “@” denote a manner of combining a shape representation with an orientation representation, and letting “o” denote a manner of combining figural representations into a representation of a cohesive group, we can visualize the structures of our representations of the configurations (a) and (b) in Figure 2.8 with the following trees:

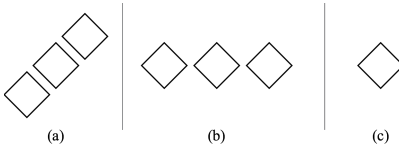
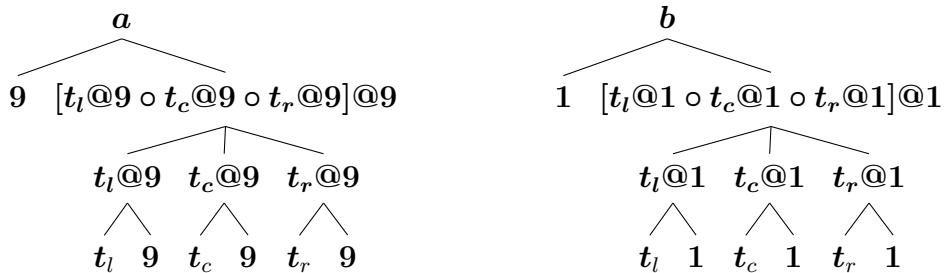


Figure 2.9: Context effects in shape perception: We perceive the very same figure as either a square or a diamond, depending on how it is situated within an array of other figures.



Attneave pointed out that the same sorts of laws governing context effects in our perception of orientation also seem to govern our perception of shape. Consider Figure 2.9. We tend to perceive the central figure f in (a) as a tilted square, whereas we tend to perceive the central figure in (b) as a vertically oriented diamond. We tend to perceive the figure in (c) as a diamond, but we can also perceive it as a tilted square. The configural biases controlling our perception of these figures are the same as those governing our perception of the triangle's orientation. It is natural to theorize that the context effects in shape are derived through some intimate relation between shape representation and orientation representation.

I will present some evidence that will clarify how shape representations and orientations are structurally related.²¹ The first point is that the representation as of a square and the representation as of a diamond are different in kind and which of these representations occurs is a

²¹The following five paragraphs repeat the discussion, in Chapter 1, of the structural relationship between shape and orientation representations. The discussion is slightly modified from its first presentation, so as to connect it to the issues of semantic compositionality that are the immediate concern here.

function of how the figure is perceived to be oriented. So orientation and shape representations seem to be structurally related. Useful evidence for this comes from priming studies.

Humphreys and Quinlan (1988) looked at how a display consisting of a square primes a display consisting of a 45° rotation of the square. A square is perceived as a square or as a diamond depending on whether we perceive its axis of orientation to bisect its sides or its corners. We have a preference for perceiving vertical axes of orientation. So if a square is presented with one of its sides parallel to the floor, we will tend to perceive it as a square. If that very same square is rotated 45°, so that one of its corners points to the floor, then we will tend to perceive it as a diamond. Humphreys and Quinlan demonstrated that there was no priming effect between perception of the square as a square and perception of the square as a diamond. This suggests that structurally distinct kinds of representations are involved in perceiving the square as a square and in perceiving it as a diamond.

The second point is that the difference in perceived orientation relative the scene or viewer does not by itself constitute the difference between the representations of the square and diamond. The lack of priming between representations as of squares and representations as of diamonds was not due merely to a difference in perceived orientation. Representations of an unambiguous figure (an isosceles triangle) at different orientations did prime each other. Indeed, even when we perceive an equilateral triangle as pointing in different directions at different times, we do not correspondingly perceive the equilateral triangle as having different shapes. And in fact there are multiple orientations that we can attribute to a square while still perceiving it as a square—we may perceive it as right-side up or upside down or on its right side or on its left side. The important thing is that whatever axis of orientation we attribute

to the square must bisect the side of the square and not its corners. That we can perceive a shape as that shape at different orientations suggests that whatever state represents as of that shape (a square, a diamond, an equilateral triangle, or an isosceles triangle) does not have a representation as of orientation (relative to the scene or to the viewer) as a constituent. This second point suggests that representations of orientation (relative to a scene or viewer) are not constituents of shape representations, but rather combine with shape representations.

What, then, accounts for the structural difference in shape representations, and why do the shape representations depend on orientation representations of the latter are not constituent parts of the former? There must be some constraint on how the two types of representations can combine. The “reference frame theory” of shape perception explains this constraint structurally (see Palmer, 1989). According to this theory—or really, this family of theories—we perceive shapes in terms of how the parts of those shapes are located and related with respect to an object-based reference frame. We can roughly think of shape representations as specifying coordinates or qualitative locations (such as *top*) for parts the shape as well as relations like *above*, *parallel*, and so on (Figure 2.10). To represent a square is, among other things, to represent a horizontal line at the top of the figure, a horizontal line at the bottom, and vertical lines to the side. An orientation representation anchors the *up* of a shape representation’s reference frame to some direction in the scene (or some direction based on the viewer). The semantic rule of composition here is that a combination of a shape representation and an orientation representation represents as of that shape at that orientation—or, rather, as of the shape’s *top* (and so on) at that orientation.

We might model oriented-shape representations as having structures like these:

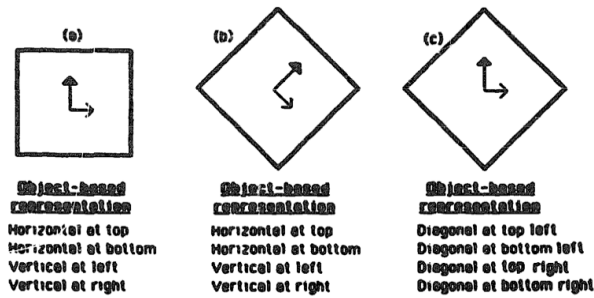
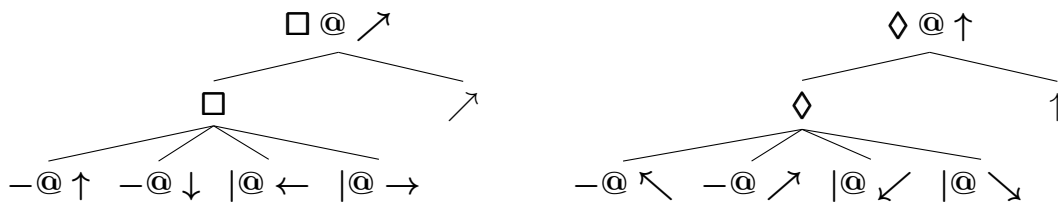


Figure 2.10: An illustration of a reference frame-based shape representation. (From Hinton and Parsons, 1988, p. 2).



Only certain combinations of orientation and shape representations are structurally possible. Roughly, the represented orientation must line up with the object-relative up-down axis of the represented shape. If the perceived orientation, together with the retinal input, implies that a corner rather than a horizontal side of the square is at the top of the shape, then that orientation representation must be combined with a representation as of a diamond. According to a reference frame theory of shape perception, then, the structure of a shape representation is constrained by the orientation representation with which it is combined. The context effect in orientation perception infiltrates shape perception because the possible structure of a shape representation will depend on the orientation representation with which it is combined.

This account illustrates how context effects, far from violating compositionality, may be the results of quite intricate structural interdependencies between perceptual representations. In the first place, representations of the orientations of items within a configuration are

structurally interdependent. This is because, for instance, we are biased toward combining representations of items as at parallel orientations. In the second place, shape percepts and orientation percepts are structurally interdependent. Plausibly, this is because shapes are represented in terms of the positions of their parts within an object-based reference frame. Given a retinal input, there are only certain ways of assigning object-based positions to the parts of a shape that are consistent with our representation of that shape's orientation.

2.5 Compositionality as a Research Program

I have offered some general arguments for the claim that perception is semantically compositional, and I have looked at how to handle three sorts of phenomena that may seem to place pressure on that claim. I argued that compositionality is compatible with and valuable in spite of the fact that the order of perceptual processing does not reflect the putative order of semantic composition. Further, I argued that emergent contents do not challenge compositionality but they do indicate that the semantic rules of composition will often be quite complex. Finally, I sketched different ways in which we can treat perceptual context effects compositionally and I indicated some of the empirical consequences of those different treatments.

I have had two aims here. The first has been to demonstrate that there is no principled obstacle to treating perception compositionally. The second aim has been to recommend, in part through illustration, the formulation and testing of compositional semantic theories of perception. Even if one is persuaded that perception is compositional, it is not at all trivial to say *how* perception is compositional. As we have seen, the theoretical alternatives are many

and subtle, and they require more targeted development and empirical evaluation. I have taken the time to explore several of these alternatives in order to illustrate that the work is non-trivial, that there are genuine puzzles to be worked out, and that empirical evidence can be brought to bear in evaluating the different alternatives.

In the next chapters, I will argue that by more closely investigating the compositional semantics of perceptual phenomena, we can get better purchase on longstanding questions in the philosophy of mind—including the question of how perception differs from thought (Chapter 3), and the question of how to reconcile the objectivity of perception with its perspectival character (Chapter 4). I would like to close, here, by recommending that perceptual psychologists have much to gain in adopting the program of producing explicit, formal compositional accounts of perceptual phenomena.

For the most part, the preceding discussion has not suggested any major upheavals to existing psychological theories. What we have done, rather, is re-frame those theories from the perspective of providing a compositional semantics. One virtue of doing this is that we have a more explicit framework for articulating some foundational assumptions and aims of perceptual psychology. For example, compositionality codifies the common practice of pairing structural descriptions with semantic descriptions. It provides an explicit framework for understanding how the perspectives provided by some perceptual states can be understood in terms of the relationships between other perceptual states.

A compositional theory must explicitly regiment our descriptions of perceptual representations into descriptions of (a) how perceptual representations are structured, (b) what they

represent, and (c) how what they represent is a function of what their constituents represent and how they are structured from those constituents. And we must distinguish these descriptions from descriptions of (d) how perceptual representations are processed. Perceptual psychology pursues each of these types of descriptions, but does not always regiment them. Such regimentation would make explicit alternative theories that might not otherwise be distinguished. It motivates us to pursue the empirical and explanatory consequences of those theories in a targeted way. Compositional semantics provides a paradigm for identifying puzzles—hard cases, like context effects, that challenge us to refine our theory of the structure and semantics of perception. It provides a framework for carving out potential solutions, which can then be pursued empirically.

Psychology is heavily invested in theorizing about the semantics of perceptual states. A great deal of psychophysical investigation is concerned with what it is that we can perceptually represent. So why has psychology largely neglected to develop compositional semantic theories? I suspect one reason is the prominence that psychology accords to accounts of *processing* or *causal relations* between representations. We have seen that compositional semantic relationships between perceptual states may not mirror the causal relationships between those states. One might think, then, that there is no explanatory value in studying those semantic relationships. However, I have argued that explicit compositional theories of the semantics of perceptual states can illuminate the causal relations between perceptual states and help us to better understand the nature of those states.

Philosophers often over-emphasize the importance of analogies between perception and language. But I believe this is a case where psychologists have under-emphasized the analogy

to the study of language. There is a fertile research program to be had in developing explicit compositional theories of perceptual phenomena, forcing us to clarify fundamental concepts in our theories of perception, and allowing us to better understand the nature of our perceptual states and their interrelations.

CHAPTER 3

The Semantic Significance of Perceptual Structures

[I]t would be a mistake to try to render explicit the semantical potential lodged in the construction of stacking modifiers by saying that our original sentence has the same semantic structure as some sentence containing the connective ‘and’. For such a sentence achieves what is admittedly the same net effect by means of different semantic elements and different constructions.

Gareth Evans

SEMANTIC STRUCTURE AND LOGICAL FORM

In the previous chapter, I defended the principle that the content of a perceptual state depends on two factors: (a) the contents of the state’s constituents, and (b) how the state is combined from those constituents—the representation’s *structural form*. In this chapter, I will argue that the structural forms of visual states contribute content about the spatial arrangements of the things that those states represent.

Different systems of representations may trade off in how much semantic material is derived from the parts of representations in that system and how much is derived from the structural forms of those representations. In classical logic, for example, structural form provides

only a very general, abstract semantic scaffolding. Consider a subject-predicate sentence, such as $Runs(usain)$, in a classical first-order language. Suppose that $Runs()$ denotes the property of running (or the set of things that run, if you prefer), $Smiles()$ denotes the property of smiling (or the set of things that smile), $usain$ refers to Usain Bolt, and $obama$ refers to Michelle Obama. That $Runs(usain)$ is about Usain Bolt's running, rather than Michelle Obama's smiling, depends entirely on what the parts of the sentence are. The sentence, $Smiles(obama)$, has the same subject-predicate form. However, because this latter sentence has different parts, it is about something very different—it is about Michelle Obama's smiling, rather than Usain Bolt's running. In these examples, the *form* of the representation, $\lceil F(a) \rceil$, expresses only something very general and abstract: that whatever term fills in for F denotes a property (or set), whatever term fills in for a refers to an individual, and that this individual has the property (or is a member of the set) denoted by F . It is the specific parts that tell you which individual and which property are being represented.

I will argue that structural form contributes much richer semantic material in the case of perceptual representation. In particular, I will argue that the very way that certain human visual states are structured commits those states to being about specific kinds of spatial arrangements. The claim is not just that we perceive things in the world as spatially arranged in certain ways. Rather, the claim is that our representation of those spatial arrangements is embodied in the very structural forms of our visual states, even abstracting away from what the specific parts of those states are.

While I will focus here on only a few components of human spatial vision, I will offer some reasons to think that analogous points can be made for perceptual states across modal-

ities and across animals, with respect to content about spatial, temporal, color, and perhaps other types of arrangements. I conjecture that in general the structural forms of perceptual states carry content about ecologically significant kinds of relations—relations whose representation plays a central role either in the survival of the perceiver or in the perceiver’s accurate representation of other important features of its environment.

I will suggest that we can elucidate central differences between types of mental capacities by examining the semantic significance of the structural forms that characterize those types of mental capacities. For example, it is plausible that the structural forms of human thoughts have much more abstract significance than those of perception. This difference reflects the distinct functions of perception and human thought. Whereas perceptual capacities function to represent quite specific aspects of one’s immediate and evolutionarily relevant physical environment, human thought functions to integrate content about potentially disparate domains. Thoughts cannot be constrained by their form to being about specific types of spatial, temporal, or other such arrangements.

In Section 3.1, I introduce the notion of *structural entailment* in order to isolate the semantic significance embedded in the structural form of a representation. In Section 3.2, I consider some preliminary reasons to think that many visual representations have *array-like structures*, in which the represented positions of things are reflected in the structures of the representations. The psychologist Zenon Pylyshyn has raised questions about the force and scope of these considerations. I intend to show that many types of visual structural form—even those that are not themselves array-like—carry spatial content. In Section 3.3 and Section 3.4, I provide arguments to show that two pervasive types of visual structural form—what I call

feature binding and *configural binding*—carry rich content about how things are spatially related. In Section 3.5, I argue that one can best explain how feature binding and configural binding operate by supposing that they depend on an underlying array-like structure. The arguments in Sections 3.3, 3.4, and 3.5 should help to address some of Pylyshyn’s criticisms. In Section 3.6, I provide some reasons to think that in general, across modalities and across animals, the structural forms of perceptual states will carry specialized content about ecologically important types of patterns (spatial, temporal, and otherwise). I conclude, in Section 3.7, by suggesting that one can better isolate central differences in mental capacities by examining the ways in which the structural forms that characterize those capacities contribute different kinds of content about the world.

3.1 What Structure Alone Contributes

Perceptual states can be complex—they can have other perceptual states as parts. My current perceptual state is a synthesis of my state of perceiving the tabletop’s *brownness* and my state of perceiving its *rectangularity*. These states are themselves complex—made up of further representational sub-states, which are not so easy to identify without the help of carefully controlled psychophysical experiments and nuanced theory. It may seem mysterious what it means to say that a mental state is structured. It is somewhat clear what it means to say that *brain states* are structured: a pattern of neurons spiking can be decomposed into events of individual neurons firing. When one says that mental states can be structured, the relevant concept of structure is more abstract and more theoretical than that of familiar spatial and

temporal kinds of structure. Like the syntactic structure of a sentence, the structure of a mental state is not something that one can directly observe, even with sophisticated instruments.

As I suggested in Chapter 1, one can best understand the claims about how mental states are structured by looking at what such claims can explain. In the first place, it is now standard in psychology to explain mental capacities as the products of computations over structured representations. Mental capacities are understood in terms of transformations over complex internal mental structures. Different views about how mental states are structured figure into different theories about how such states are generated and employed in our psychologies.

In the second place, hypotheses about how mental states are structured help to explain which mental states can and cannot occur in individuals of a given species, and to explain systematic interdependencies between mental states. The notion of structure comes along with a notion of structural possibility—whether a certain set of representations can, cannot, or must be combined. One way to explain why one cannot perceive a color as yellowish blue or reddish green is that such a representation is not *structurally* possible (Hering, 1964; Hurvich and Jameson, 1957). (By contrast, some psychological states are structurally possible, but one never encounters the right stimulus, or intervening factors upset the production of the state, or one lacks the general psychological resources necessary for undergoing that state.)

It is unclear how one could come to perceive so many novel shapes over the course of one's life—think of any new piece of Ikea furniture or any Frank Gehry building—except by having some stock of basic types of perceptual states that can combine and re-combine into new, complex perceptual states. It is also difficult to explain systematic relationships among

one's perceptual capacities without positing that one's perceptual states can share constituent parts (compare Frege, 1963; Fodor, 1987; McLaughlin, 2009). It is a law of human psychology that if one can perceive a scene as containing a red circle besides a blue square, then one can also perceive a scene as containing a red square besides a blue circle. It is plausible that the reason one's capacity to perceptually represent a red circle besides a blue square necessitates a capacity to perceptually represent a red square besides a blue circle is that one has more basic abilities to perceive redness, blueness, circularity, and squareness, and that these can be combined and recombined in different ways, depending on the stimulus conditions and one's perceptual set (that is, one's expectations, preferences, memories, and so on). A good deal of perceptual psychology has been dedicated to working out the ways in which perceptual states are organized from more basic perceptual states and what those more basic perceptual states are.

There are different ways in which perceptual states can be combined—different modes of combination. Perceptual modes of combination each correspond to compositional principles governing how the combination's denotational content—what it purports to denote—relates to the denotational contents of the combination's constituents. One might posit that my representation of the tabletop in front of me as brown and rectangular is a *feature-bound* construction from a representation as of an instance of *brownness* and a representation as of an instance of *rectangularity*. Feature-bound constructions, as a rule, represent all the properties denoted by the constituent perceptual states as co-instantiated in a particular individual (Treisman and Schmidt, 1982). In addition, one might posit that there are *configurally bound* constructions from representations of particular individuals in a scene. Configurally bound constructions

represent the set of elements denoted by their constituents as a *cohesive group*. For example, my representation as of a group of several animals is a configurally bound combination of my representations of each individual animal (Palmer, 1977).

Each of these putative modes of combination is defined over certain structural categories of representations. The *structural category* of a representation determines which other representations it can combine with. By analogy, first-order logic contains categories such as *n-ary predicate* and *constant*. Constants cannot be combined merely with other constants, but must be combined under the scope of a predicate. An *n-ary* predicate can be combined with *n* constants, but not *n + 1* constants. It is a non-trivial empirical project to specify the structural categories of a naturally occurring representational system. In English, there are categories such as *noun*, *verb*, *auxiliary*, and *preposition*, and there are rules governing how expressions of these categories can be put together. Plausible candidates for structural categories of perceptual states include: *feature representation* (including, as sub-categories, *color representation*, *shape representation*, and so on), which can combine into *feature-bound representations*, *object representation*, which can combine into *configural representation*, and so on. A representation's *structural form* is determined by the structural categories of its constituents and their mode of combination.¹

I will be arguing that we should countenance such perceptual modes of combination as

¹I will for the most part concentrate on how representations are combined from their constituents, putting to the side interesting questions about how to demarcate the structural categories of those constituents.

One may wish to say that a representation's structural form is in some sense a "part" of that representation. This is okay if one clearly distinguishes that some parts of a representation have a designated formal role in governing how the other parts are combined. I prefer to say that a representation's structural form is an *aspect* of the representation.

feature binding and configural binding, that they cannot be reduced to more familiar types of logical constructions such as conjunction and predication, and that these modes of combination carry substantive content about how things are spatially arranged. There are very likely many other perceptual modes of combination, about which the analogous points can be made.

The content of a representation depends on its structural form (and associated compositional rules) and on the contents of its basic parts. I will adopt the notion of *structural entailment* to capture the semantic contribution made just by the form of a representation, abstracting from the contributions made by the specific constituents of that representation (cf. Evans, 1985b; Balcerak Jackson, 2017). On a first pass, a representation structurally entails a proposition p just in case that representation's mode of combination and the structural categories of its parts dictate that the representation cannot be veridical unless p is true. It will be useful to adopt a more general notion of structural entailment, where a representation structurally entails not just propositions but also that certain functions from individuals to propositions (*propositional functions*) are true of (*satisfied by*) relevant sequences of individuals.²

²Relatedly, Evans (1985b) discusses inferences that he calls *structurally valid*. Evans offered a reductive analysis of structurally valid inferences, leaning heavily on Tarski (1956)'s analysis of logical consequence: "an inference from $S_1 \dots S_{n-1}$ to S_n is structurally valid iff... for every interpretation, every sequence which satisfies $S_1 \dots S_{n-1}$ upon that interpretation also satisfies S_n upon that interpretation" (Evans, 1985b, p. 71). Admissible interpretations will be constrained by the syntactic analyses of these expressions: different syntactic categories are to be assigned different kinds of semantic values, and how an expression is combined from its parts will determine how its semantic value is a function of the semantic values of its parts. As Evans emphasizes, the notion of structurally valid inference, unlike the notion of logically valid inference, does not rely on selecting a set of logical constituents whose interpretations are to be held fixed.

I rely here on an intuitive understanding of entailment and of entailment-in-virtue-of-structure. One might wish to adapt Evans' definition of structurally valid inference to analyze this intuitive notion. But while a reductive analysis of structural entailment along these lines might be possible, Evans' analysis is not entirely satisfactory. For one, it is not clear that the analysis can account for the modality involved in our intuitive notion of entailment. The material fact that all models of $S_1 \dots S_{n-1}$ are models of S_n may not be able to account for the modal sense in which $S_1 \dots S_{n-1}$ *cannot* be true (*/satisfied*) unless S_n is true (*/satisfied*). Likewise, the question of whether a Tarski-style definition of logical consequence can capture the modal force that is attached to the intuitive notion of consequence has been amply debated in philosophy of logic (Etchemendy, 1990; Sher, 1996; Shapiro, 2005). Additionally, Evans' definition of structural validity, much like Tarskian accounts of logical consequence, runs

To apply the notion of structural entailment, one must have a theory of how the representations in question are structured. By contrast, accounts of *logical entailment* are typically burdened by the need to specify a set of *logical constituents* in virtue of which a proposition entails certain things. It has proven extraordinarily difficult to come up with a motivated demarcation of logical constants (see, for example, Tarski and Corcoran, 1986; Peacocke, 1976; Sher, 1996; Bonnay, 2008). However, it is a different sort of project, and an empirical one at that, to identify the modes of combination that characterize a naturally occurring representational system.

I appeal to structural *entailment* rather than structural *representation*, in order to capture the entire semantic import of a representation's structural form. A representation may structurally entail that certain propositional functions are satisfied by relevant elements, even if those propositional functions are not explicitly specified or expressed by the representation. For example, a drawing of a cube in linear perspective may structurally entail that the cube projects a certain pattern onto the picture plane, without *representing* that cube as doing so.³

Relatedly, it may be that a representation has structural entailments that are not specifically represented by *any* representations in that system. One standardly thinks of entailment as a relation among representations in a system. However, I do not assume that perceptual sys-

into issues with providing a complete, principled specification of the admissible interpretations (for the analogous discussions about logical entailment, see Varzi, 2002; Beall and Restall, 2006). A closely related issue is that where p is a structural consequence of a representation, one may not want to also count as structural consequences logically equivalent propositions, such as $p \wedge (q \rightarrow q)$, or necessary propositions, such as $2 + 2 = 4$. Because of these sorts of issues, I stick to intuitive notions of entailment and entailment-in-virtue-of-structure, rather than trying to offer a reductive definition.

³By analogy, I argue, in Chapter 4, that one does not perceptually represent things as projecting particular patterns onto the retina. Yet a given perceptual representation may *structurally entail* that the represented object projects a certain pattern onto one's retina.

tems are always capable of specifically representing the structural entailments of perceptual states. It may be that a perceptual state entails that some propositional function ϕ is satisfied by a relevant set of elements, yet there is no perceptual state that functions specifically to represent ϕ or that ϕ is satisfied by those elements. Indeed, if perceptual states do not represent propositions or propositional functions at all (see, for example, Burge, 2010b), then the structural entailments of perceptual states will never be specifically represented in perception.⁴ Nevertheless, it is plausible that transitions among perceptual states will normally respect the structural entailments of those perceptual states in the following sense: if a perceptual representation \mathbf{a} structurally entails that ϕ is satisfied by the relevant items, then, all things equal, the visual system will not transition from \mathbf{a} to a perceptual representation \mathbf{b} that structurally entails that ϕ is not satisfied by those elements.⁵

Before turning to perception, it will be helpful to see how the notion of structural entailment applies to some public representations. It is natural to think that the map in Figure 3.1 structurally entails that the landmark denoted by the icon labeled C is located to the west of the landmark denoted by the icon labeled B . In particular, left-right relations among parts of the map seem to structurally entail west-east relations among the places denoted by those parts. Abstracting away from what the particular icons in the map are or what they denote,

⁴This is another point at which the notion of structural entailment comes apart from Evans' notion of structurally valid inference (see Footnote 2). A representation may structurally entail a proposition, even when the proposition cannot be expressed in that representational system and, as a result, no inference to a representation of the proposition can be carried out.

⁵I largely follow the notational conventions from Chapter 3. Bolded letters, such as " \mathbf{a} ," refer to mental representations; lower-case, italicized, non-bolded letters that scope over other letters, such as " f " in " $f(\mathbf{a}, \mathbf{b})$," refer to structural operators or modes of combination for mental states. I follow familiar conventions for first-order logic by also using italicized, non-bolded letters to quantify over propositions (" p ", " q ," and so on) and terms (upper-case for predicates and relations and lower case for constants, as in " $F(\mathbf{a})$ ").

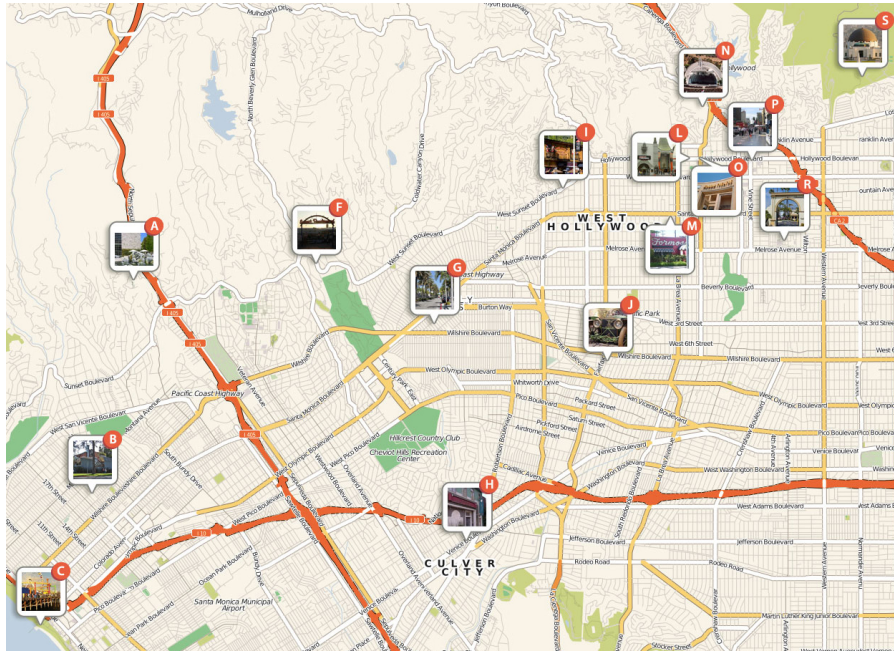


Figure 3.1: A conventional map of part of Los Angeles. (Adapted from <https://travel.sygic.com/lp/Los-Angeles-Printable-Tourist-Map>).

the map will be accurate only if whatever is denoted by an icon on the left is west of whatever is denoted by an icon on the right (see Camp, 2007; Rescorla, 2009). Likewise, if one part of a map contains another part, this structurally entails that the region represented by the former contains the region represented by the latter (Casati and Varzi, 1999). So, on the face of it, conventional maps structurally entail that certain spatial and topological relations obtain among the places denoted by the parts of the map.

Sentential representations may also have spatial structural entailments. In natural language, there is reason to think that the syntactic structures of certain prepositional phrases structurally entail certain spatial properties or relations. Spatial prepositions roughly divide into two grammatical categories: *place* prepositions, such as “in,” “at,” and “behind,” and *path* prepositions, such as “toward” and “from” (Jackendoff, 1983). These grammatical cat-

egories are characterized by different distributional patterns. For example, place prepositions can serve as arguments for stative verbs such as “be” and “stay,” while path prepositions cannot (compare: “Immanuel stays in Königsberg” with **“Immanuel stays toward Königsberg”*). While particular place prepositions each have different meanings (“under” is clearly different from “over”), each of them represents some fixed spatial location or other relative to the object denoted by the determiner phrase that serves as an argument to the preposition (for example, “Königsberg” in “in Königsberg”). Since place prepositions form a genuine grammatical category, it is an aspect of the syntactic form of a sentence that it contains a place preposition. Since place prepositions all represent spatial locations relative to the denotation of their adjacent determiner phrase, sentences that contain place prepositions structurally entail that the subject of the clause containing the place preposition (“Immanuel,” in this case) denotes something that has a spatial location relative the place preposition’s adjacent determiner phrase (“Königsberg,” in this case).⁶

Other representations have much more abstract structural entailments. The Euler diagram in Figure 3.2 structurally entails that the set denoted by the smaller circle, labeled *A*, is a subset of the set denoted by the larger circle, labeled *B* (see Shin, 1994). The Euler diagram does not structurally entail anything about the spatial arrangement of its subject matter. Likewise, sentences of the form $\ulcorner F(a) \urcorner$ in classical first-order predicate logic, such as *Runs(usain)* and *Smiles(obama)*, structurally entail that the name that fills in for *a* refers to an individ-

⁶For more detailed discussion of the syntax and semantics of spatial prepositions, see (Kracht, 2008; Koopman, 2010; Svenonius, 2010).

I suggested earlier that it is a feature of conventional maps that they structurally entail spatial relations. However, if some natural language sentences also structurally entail spatial relations, then having spatial structural entailments is not a sufficient condition for being map-like as opposed to sentence-like.

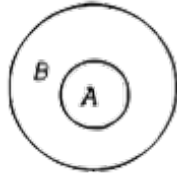


Figure 3.2: An Euler diagram, depicting A as a subset of B .

ual, that the predicate that fills in for F denotes a property, and that the named individual instantiates the denoted property. Sentences of this form do not structurally entail any spatial arrangements. The first-order sentence $Westof(missoula, bozeman)$ does not *structurally* entail that Missoula is west of Bozeman. While the sentence lexically entails that Missoula is west of Bozeman, other representations of the same form (predicating a binary relation of two individuals) need not entail anything about Missoula, Bozeman, or their spatial relationship.

I have introduced the notion of structural entailment to isolate the semantic contributions made by just the structural form of a representation. I will now argue that certain types of complex visual states structurally entail that specific kinds of spatial relations obtain among the things represented by those states. I will focus on vision, and in particular on three different structural forms of visual state: (a) array-like structures, (b) feature-bound structures, and (c) configurally bound structures. I expect that arguments that I use in each of these cases can be extended to other types of visual perceptual states, to non-visual perceptual states, and to perceptual states in other animals.

3.2 Visual Arrays

Many philosophers and psychologists have suggested that a central class of visual representations have *map-like* or *array-like* structures (Marr, 1982; Kosslyn, 1980; Pinker, 1988; Tye, 1991; Peacocke, 1992; Burge, 2014). The thought is that the structural form of these representations reflect the structure of the visual field—the set of directions in which we locate things. I will show how array-like representations structurally entail spatial relations among the things they represent. I will then provide some initial reasons to think that many visual states are array-like. By the end of the paper, I will have given further arguments for the centrality of array-like representations in vision, while also showing that other types of visual structural form have spatial significance. In fact, I will suggest that these other structural forms are best understood as incorporating array-like representations.

To say that certain visual states are array-like is to claim that certain visual states occupy cells, or have addresses, in an abstract array, where each cell or address corresponds to some aspect of the spatial position in which the represented item is perceived as located. For example, it may be that cells in the array determine the directions in which one perceives things, while the representations that occupy those cells determine *what* one perceives (for example, a black dot) and how far away one perceives it to be (for example, Figure 3.3). Visual states can be *combined* by virtue of each having an address in the array. The relations among cells in the array are isomorphic to relations among visual directions in the following way. If two visual states occupy adjacent cells in the representational array, then the whole array-like representation is accurate only if the things represented by those component states lie in adjacent

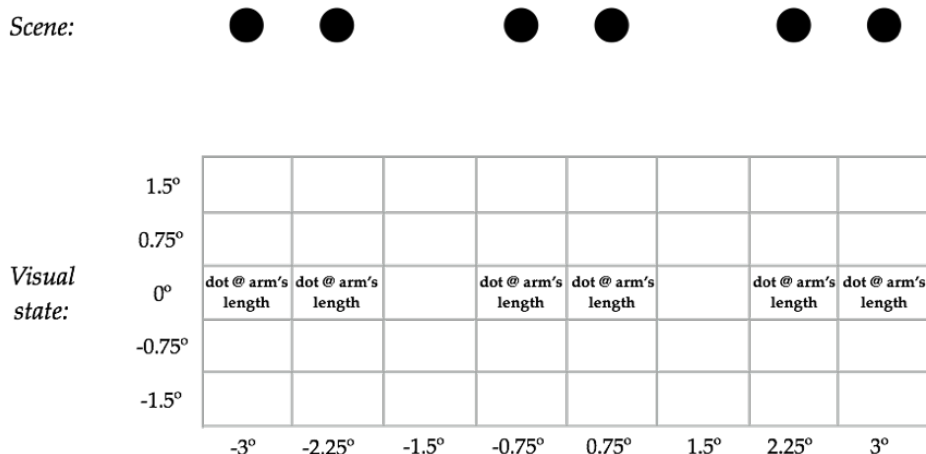


Figure 3.3: An illustration of a simple scene and the structure of an array-like visual representation of that scene. Cells in the array correspond to directions from the perceiver’s viewpoint (indicated here in degrees of angle in the horizontal and vertical directions). The representations occupying those cells represent features, such as shape, as at certain distances from one’s viewpoint.

directions from the perceiver’s viewpoint.⁷

By definition, a representation’s cell or address in the representational array is a structural feature of the representation.⁸ The computations and combinations into which a representation can enter depend on what cell it occupies in the array. To say that a perceptual representation is array-like is not to say it is literally laid out in space, nor does it suggest that we have

⁷Following Marr, I am focusing on two-dimensional array-like representations, in which cells correspond to lines of sight from the viewer. In such a scheme, the distance to an item in the world is represented by a constituent of a representation, not by the structural form of the representation. There may be other types of array-like representation. An array-like representation could be three-dimensional and viewer-centered, with cells corresponding to both line of sight and distance from the subject’s viewpoint, or three-dimensional and object-centered, with cells corresponding to height, width, and depth relative to some point on the object. Alternatively, cells in the array could have both viewer-centered and world-centered addresses (Pinker, 1988). Elsewhere, I argue that there are object-centered array-like representations in vision that can be (and normally are) embedded within viewer-centered array-like representations (Lande, unpublished).

⁸One can think of a cell or address in the array as a structural operator, $a_{i,j}$, that takes a primitive representation \mathbf{b} of a property instance, for instance, and produces a construction $a_{i,j}(\mathbf{b})$, that is accurate only if that property instance is located along a particular line of sight, indexed by i, j . The whole representational array can be thought of as a structural mode of combination a that takes a set of representations $\mathbf{b}_1, \dots, \mathbf{b}_n$ and returns a combination of representations $a_{i,j}(\mathbf{b}_1), \dots, a_{k,l}(\mathbf{b}_n)$.

some “inner eye” that views and interprets the patterns in an “inner array” (see Block, 1983).⁹ In fact, for all I have said, array-like representations could be physically realized by linear strings of physical items, each of which carries some marker of their two-dimensional (or three-dimensional) “address” in the abstract array. The central point is that the way certain visual representations are processed and the ways they can combine with each other depend on what cells in an abstract array those representations occupy, where cells correspond to the positions or lines of sight in which things are perceived and where relations among the cells correspond to relations among those lines of sight.

A visual representation that is addressed by a particular cell will, by virtue of having that address, structurally entail that referent of the representation is located along a particular line of sight from the viewer. Indeed, any representation that has that address will be accurate only if its referent is located along that same line of sight. An array-like representation that consists of several addressed representations of items will structurally entail that those items subtend a particular visual angle—that those items will subtend an angle between the lines of sight corresponding to the addresses of the relevant representations. Array-like perceptual representations will also structurally entail certain topological relationships: if a representation occupies a cell while a set of representations occupy surrounding cells, this structurally entails that the item denoted by the former representation is along a line of sight surrounded by the items denoted by the latter representations.

Computational theories in psychology offer one basis for thinking that a large class of

⁹It is noteworthy, however, that areas of the brain dedicated to visual processing are laid out retinotopically, so that adjacent parts of the brain are dedicated to processing information from adjacent parts of the retina (Frisby and Stone, 2010, Ch. 10).

visual representations are array-like. Computational theories explain how subjects come to perceptually represent aspects of the world, such as shape, size, and color, on the basis of proximal information available to their sensory receptors (Marr, 1982; Palmer and Kimchi, 1986). How, for example, does the visual system generate a representation of a surface's shape and location in three-dimensions on the basis of a two-dimensional retinal image? The central insight of computational psychology is that one can treat this problem as one of decomposing the capacity into a series of more basic, interlocking transformations. In order to characterize these transformations with mathematical precision, the psychologist must specify the types of structure that enter into the domain and range of each transformation.

An overwhelming number of theories in perceptual psychology posit that outputs of key transformations in the visual system are array-like representations. Summarizing the psychological literature on shape perception, for example, the psychologist James Todd writes:

Almost all existing theoretical models for computing the 3D structures of arbitrary surfaces from visual information are designed to generate a particular form of data structure that can be referred to generically as a 'local property map'. The basic idea is quite simple and powerful. A visual scene is broken up into a matrix of small local neighborhoods, each of which is characterized by a number (or a set of numbers) to represent some particular local aspect of 3D structure, such as depth or orientation. (Todd, 2004, p. 118).

Witkin (1981), for instance, describes a model for representing the orientations of surfaces in depth, relative to the viewer. The output of Witkin's model is an array of surface ori-

entation representations, each of which is indexed by a line of sight from the viewer. (Strictly speaking, in Witkin's model the representations of surface orientation are indexed by the position of their projection on the retinal image, which corresponds to a line of sight to the surface.) Theories of color vision also rely on array-like structures. What color one perceives a surface to have depends on one's perception of neighboring surfaces. Retinex theory was one of the first theories to recognize the importance, in color perception, of computing the relationships between luminance values at different points of the eye (Land, 1977, 1983). These computations require different luminance values to be indexed and ordered by their relative positions on the eye. The outputs of these computations are representations of color that are likewise indexed according to the line of sight in which the color perceived (see also Horn, 1974).

There is good reason that these computations should operate over array-like structures. Burge (2014, p. 494) remarks that

The format of visual representation takes on some of the geometry of the mapping of light on the retina. Representation via ego-centrally anchored spatial coordinate systems [that is, the array-like structure of visual representations] owes much to the spatial layout of light registration by retinal receptors. Some of this layout of pre-perceptual registration of retinal information is preserved and co-opted by the perception-formation process.

The retinal images that provide the main inputs to vision are spatial arrays of light intensities on the retina. Spatial relationships within those arrays of light intensities provides significant information about the distal sources of those light intensities. By comparing neigh-

boring light intensities, the visual system can begin to determine where there are edges in the distal environment (Marr and Hildreth, 1980). There are sophisticated mathematical tools for describing transformations over arrays. These mathematical tools offer elegant descriptions of how an array that encodes light intensities at different points on the eye can be transformed into an array that represents distal features, such as surface orientation and color, at different locations the world. It is natural that retinal images should be encoded as an informational array, in which adjacent light intensities are registered in adjacent cells in the array, and that analyses over that array should have the form of transformations that output array-like representations of the world.

However, there are some limits to these computational arguments for array-like representations in vision. In the first place, one could argue that the computational arguments are not really committed to specifically array-like structures, but merely to representations that carry content about how things are arranged in space with respect to a viewer. Pylyshyn has argued in this way against the view that visual mental images—the representations that underlie our capacity to visually imagine scenes—are array-like. Standard arguments that mental imagery is array-like cite evidence that in order to mentally “scan” from one point in an imagined scene to another point, one seemingly has to scan intermediate points in the imagined scene. Kosslyn (1973) argues that this is best explained by supposing that mental images are array-like in structure, and that scanning an imagined scene involves an operation that traverses through the cells of such a representation. Pylyshyn (2002, p. 167) critiques such processing arguments, writing that

there is nothing to prevent us from modeling the content of an image as a set of sentence-like expressions in a language of thought. We could then stipulate that in order to go from examining one place (referred to, say, by a unique name) to examining another place (also referred to by a name), you must pass through (or apply an operation to) the places whose names are located between the two names on some list. You might object that this sort of model is ad hoc. It is. But it is no more ad hoc than when the constraints are applied to a matrix formalism. Notice, moreover, that both become completely principled if they are taken to be simulations of a real spatial display.

Pylyshyn's point is that what appear to be computations that operate on array-like structures could just as well be computations that operate on subject-predicate representations, which relate particular features to particular locations in the scene.¹⁰ One could argue that there is no principled reason that mental scanning would be constrained to operate on such representations in order according to the locations they attribute to items in the world. But Pylyshyn argues that there is no more or less reason that mental scanning should be constrained to traverse an array from one cell to another, crossing each intervening cell, as opposed to hopping between the distant cells.

Pylyshyn's argument does not seem to have moved the majority of psychologists, and they have had little impact in the domain of perception. It is striking how many computational

¹⁰One could characterize such representations as having the form $Loc(c, (a, b))$, where c represents a feature such as surface orientation or color, (a, b) represents a particular direction from the perceiver's viewpoint, and Loc denotes the relation *is located at*.

models in psychology appeal to array-like representations of space rather than the subject-predicate spatial representations that Pylyshyn offers as an alternative. One would do well to take the near ubiquity of array-like posits in successful, elegant explanations as a compelling consideration in favor of the reality of those posits. It is very natural that many visual states should be array-like, given the structure of the retinal image and the availability of elegant mathematical formalisms for describing how the retinal image can be transformed into an array-like representation of the distal scene. Yet Pylyshyn's arguments do challenge one to isolate the specific empirical differences between an array-like representation of space and a representation whose constituent parts, rather than form, denote spatial positions and relationships.¹¹ Can we cite anything more in favor of array-like posits in addition to their naturalness and sociological prevalence in perceptual psychology?

Moreover, a number of visual states are not obviously array-like. Consider feature binding, in which representations of individual feature instances combine into a representation of those two features as co-instantiated in a particular individual. One might think that feature binding, for instance, is a form either of conjunction or, perhaps better, predicate conjunction (Clark, 2004). Or consider configural binding, in which representations of individual elements in a scene are combined into a representation of a cohesive group, of which those elements are members. One might think that configural binding really just amounts to attaching a predicate, denoting *grouphood*, to the representations of the individual elements (cf. Watt, 1988; Pylyshyn, 1989; Feldman, 1999b,a). Neither conjunction, predicate conjunction, nor mere

¹¹That is, one wants to know what the empirical difference is between a representation of the form $a(c_1, \dots, c_n)$ (as illustrated in Figure 3.3) and a set of representations of the form $Loc(c_1, (a_1, b_1)), \dots, Loc(c_n, (a_n, b_n))$.

predication structurally entail anything about how things are spatially related. Even if one allows that many visual states are array-like, one might think that there are many other types of visual states the structural forms of which do not have spatial significance.

In the next sections, I will argue that these latter types of visual states do structurally entail that things stand in certain spatial relations. Spatial content is embodied in the very forms of these visual states, even abstracting away from their specific constituents. Spatial structural entailments can be found throughout vision, and are not limited to array-like states. First, I will argue that the structural form of feature-bound states dictates that they represent features as being located in the same place. Second, I will argue that the structural form of configurally bound states dictates that they represent things as being either sufficiently close to each other, or similar in shape, orientation, or size. I will then argue that these points provide further support for the claim that visual states have array-like structure, in which the spatial positions of represented items are reflected in the structural form of visual representations. Feature binding and configural binding are best seen as operating over array-like states.

In making these arguments, I hope to better isolate empirical differences between views on which spatial relations are represented by the structural form of a representation and views on which such relations are merely represented by constituent parts of those representations. I will be relying on the same general form of argument throughout. I will argue that there are certain constraints on how visual states can and cannot structurally combine with each other (see Chapter 1). Visual states can combine in certain ways only if those states represent things in such a way as to entail that those things stand in specific spatial relationships with each other. Insofar as those implied spatial relationships are built into the very constraints on

structural combination, spatial content must be an aspect of the structural forms of the states in question. I do not believe this powerful type of argument has been offered before.

3.3 Feature Binding

I see the tabletop in front of me as a brown, rectangular individual. It is very plausible that the state I am in when I perceive the tabletop as a brown, rectangular individual, has as distinct constituents my state of seeing the tabletop's brownness and my state of seeing its rectangularity. One cannot be in the former state without being in the latter states. Further, the state of perceiving the tabletop's brownness is distinct from the state of perceiving the tabletop's rectangularity. One can be in either of these states without being in the other. Indeed, there is a large amount of evidence that representations of features such as color and shape are processed relatively independently in the visual system (Garner, 1974; Treisman and Schmidt, 1982; Nissen, 1985; Cant et al., 2008; Cavina-Pratesi et al., 2010). I call my perceptual representation of the tabletop as a brown, rectangular individual a *feature-bound* combination of my representations as of instances of brownness and rectangularity.¹² As part of a compositional rule for feature binding, one may say that a feature-bound representation is accurate only if the feature or property instances denoted by its constituents are co-instantiated in a particular

¹²Treisman (1986) calls these complex representations "object files." It is an empirical question whether the "individuals" in which we perceive features to be co-instantiated correspond to the more familiar physical bodies that we intuitively consider to be "objects." The latter type of individual is characterized by, among other things, cohesiveness, boundedness, and rigidity (Spelke, 1990).

individual.¹³

Feature binding, as I have described it so far, may sound equivalent to something like predicate conjunction (see Clark, 2004, p. 451). The expression, $\lceil x \text{ is brown and rectangular, } \rceil$ involves the conjunction of the predicates “brown” and “rectangular,” and is true of a situation in which x has both the property denoted by “brown” and the property denoted by “rectangular.” However, I will argue that perceptual feature binding is not reducible to predicate conjunction or to any familiar operation from classical logic. Any feature-bound representation, no matter what its constituents are, does not merely represent the properties denoted by

¹³Many studies of feature-binding are ambiguous about whether a feature-bound state represents the relevant features as co-instantiated *in a particular location* or as co-instantiated *in a particular individual*, which may occupy different locations at different times. A number of studies suggest that feature-binding is normally of the latter sort. There may also be a type of feature binding that simply amounts to representing two features as co-instantiated in a particular place (Clark, 2004). However, I will here be focusing on the type that involves representing features as co-instantiated in the same individual.

Pylyshyn (2003) cites, as evidence for this latter sort of feature binding, a study by Blaser et al. (2000), in which subjects are shown a display consisting of two overlapping Gabor patches, each of which varied independently in appearance along various dimensions, such as color and orientation. As Pylyshyn describes it, even though subjects perceived the patches as being located in the same place, they did not join together their representations of all the features visible in that place. However, Pylyshyn misdescribes the case. Blaser et al. reported that “The Gabors appeared to observers as simultaneously present and transparently layered on one another, although there was no noticeable separation in depth” (Blaser et al., 2000, p. 196). Even if there was “no noticeable *separation* in depth,” the perception of one surface as transparent and layered on top of another suggests that one does not perceive the two surfaces as at identical depths, and so does not perceive the surfaces as at exactly the same particular location. At most, the study shows that representing two features as located along the same line of sight is not sufficient for feature binding.

Other studies offer more compelling evidence that feature binding is not merely the representation of features as co-instantiated in a particular location. For example, Kahneman et al. (1992) showed an “object-specific preview benefit.” In a typical demonstration of the object-specific preview benefit, two squares are presented on-screen, each containing a letter (or color, or what have you). The letters then disappear and the squares move to new positions. Letters then appear in the displaced squares. The subjects’ task is to indicate whether one of the initially displayed letters is present in the final display. Accurate responses are much faster when the target letter reappears in the same square in which it initially appeared, even though that square is now in a new location. The initial feature-bound representation of the square containing the target letter primed the later feature-bound representation of that square and letter at a different location. If feature binding involved merely representing features as co-instantiated in a particular location, there should not be a selective priming advantage for this condition, since the final display contains those features at a new location. The priming can be explained, however, by supposing that we represent the square and letter as co-instantiated in the same individual, which subjects can track as it moves from location to location. Subjects’ initial representation of the square and letter as features of the same individual primed their later representation of that same individual.

its constituents as co-instantiated in an individual, it structurally entails that those properties *spatially coincide*. Predicate conjunction and syntactic operations in classical logic lack such structural entailments.

The key phenomenon here is that if we represent the instance of brownness as over *here* and the instance of rectangularity as over *there*, then it seems we cannot represent the brownness here and the rectangularity there as co-instantiated in the same individual. For feature representations to be combined, the represented features must be perceived as in the same location (Treisman and Gelade, 1980; Treisman, 1986).¹⁴ I intend to leave it open whether feature binding requires attributing the same specific location, at a given time, individually to each feature or whether it requires representing both features as *in the same location* (wherever that location may be) at a given time.¹⁵ Note that the claim, that feature binding requires that we locate the relevant features in the same place at a given time, is consistent with the evidence that one can represent the features as co-instantiated in an individual that can move over time (Kahneman et al., 1992). The features, as represented, must spatially coincide at

¹⁴To be more precise, if we represent the instance of brownness as *here* with respect to a particular perceptual reference frame and if we represent the instance of rectangularity as *there* (\neq *here*) with respect to that same perceptual reference frame, then we cannot feature-bind those representations.

A central part of Treisman's theory is that feature binding requires focally attending to the locations of the represented property instances. On early models, focal attention was supposedly necessary for representing multiple features as having the same location within a common reference frame. On more recent models, focal attention is supposed to be necessary for distinguishing between the different features present at a particular location (Quinlan, 2003). I do not assume that attention is necessary for feature binding. There is a range of evidence that attention is not necessary for feature binding (Wolfe and Bennett, 1997; Thorpe et al., 1996; Rousselet et al., 2002; Potter, 2012). I am only claiming that feature-binding requires that the feature representations locate the relevant property instances in the same place.

¹⁵Pylyshyn (2003) argues that feature binding does not require representing the specific locations of the individual features at a given time. This is compatible with the latter claim that feature binding cannot occur if one perceives the features to be in different places at a given time. Still, I am skeptical of Pylyshyn's arguments. A number of studies suggest that we do not normally represent features independently of their locations, and that feature binding typically is mediated by the perceived locations of the relevant features (for a summary of some of the evidence, see Quinlan, 2003).

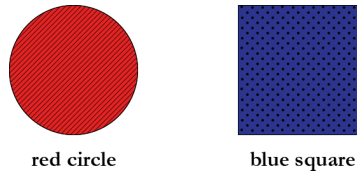


Figure 3.4: Under normal circumstances, one will perceive a red circle besides a blue square. Under certain circumstances, one may inaccurately perceive a red square beside a blue circle.

any given time. However, the features may together change place over time.

Evidence that feature binding requires that we locate the relevant features in the same place comes, in part, from examining the circumstances under which one mistakenly represents unrelated features as co-instantiated in the same individual—so-called cases of “illusory conjunction.” Under certain exceptional conditions one may perceive a display like Figure 3.4 as containing a red square besides a blue circle. Mistaken feature binding can happen under various combinations of the following circumstances: when the display is presented very briefly and then masked, when the figures are presented outside of foveal vision, when one is very distracted, or when the figures are packed into a dense display of other figures. These are all situations in which one becomes less accurate and less precise at representing the positions of the features. The farther away the two figures are, and in general the more the circumstances facilitate spatial acuity, the less likely one is to form an illusory conjunction of features from the two figures (Gallant and Garner, 1988; Ashby et al., 1996; Hazeltine et al., 1997).

Illusory conjunctions occur in proportion to the likelihood that we will see the features as spatially coinciding. All things being equal, if one misperceives two features as co-instantiated in an individual, then one has perceptually mislocated one or the other of the features. Conversely, if one has accurately and precisely located the features in different places, then, all

things equal, one will not mistakenly feature-bind one's representations of those features.

Note that it is an empirical question what exactly it is to perceive two features as *spatially coinciding*, or as being in the *same location*. The visual system likely demands different levels of precision about the locations of the features in different circumstances (Cohen and Ivry, 1991). Further, the two features need not totally overlap each other for us to perceive them as being located in the same place. When feature binding representations of the same type of feature (for example, representations of different color), as when one sees the *blackness* and the *whiteness* of a black and white cookie as co-instantiated in the same individual, spatial coincidence may amount to something like adjacency. What counts, for perception, as spatially coincident can also depend on whether the features are seen as located within a cohesive group (Prinzmetal, 1981) and on whether the features are seen to be located within what is antecedently perceived as a whole object (Naber et al., 2011).¹⁶

Psychologists and philosophers typically treat the spatial coincidence constraint on feature binding—that the represented features must be seen as spatially coincident—as either a semantic constraint on the accuracy of feature-bound representations or as a mere computational constraint on the generation of feature-bound representations (see, for example, Treisman and Gelade, 1980; Tacca, 2011). Such views are compatible with holding that feature binding does not *structurally* entail spatial coincidence. I will now argue that the spatial coincidence constraint is not merely a constraint on the accuracy of feature-bound representations and not merely a causal constraint on which feature-bound representations are generated,

¹⁶Feature binding can therefore interact in complex ways with modes of combination such as configural binding. I will not attempt to describe these interactions here.

but that it also counts as a structural constraint on which feature representations are combinable. The very combinability of feature-bound representations requires that the represented features be seen as located in the same place. If this argument is right, then feature-bound representations structurally entail spatial coincidence. Any structurally possible feature-bound representation will, just by virtue of its feature-bound structure, entail that the represented features are located in the same place.

The spatial coincidence constraint on feature binding is not merely a semantic constraint on the accuracy of feature-bound representations. One might argue that it is not the structure of a feature-bound representation that entails spatial coincidence; rather perceptual feature representations happen to represent properties, such as shape and color, that as a matter of fact can only be co-instantiated if they spatially coincide. As a result, a representation of such properties as co-instantiated could only be accurate if those properties were also spatially coincident. However, a merely semantic constraint fails to explain why representations of features can only combine if those features are represented as spatially coinciding. The merely semantic account does not have anything to say about our inability to form feature-bound representations of properties as co-instantiated yet spatially disjoint. The account implies that such a representation would be *inaccurate*, but does not suggest that they could not be formed. This leaves open the possibility that one could inaccurately perceive features as co-instantiated in the same individual even when one has correctly perceived the relevant features as spatially disjoint.

It is very unlikely—perhaps metaphysically impossible—that spatially disjoint shape and color could be co-instantiated in the same object. However, it is a distinct point that the visual

system does not feature-bind representations of features that one locates as being in different places. That a situation is metaphysically impossible does not entail that the representation of that situation is impossible, only that it is necessarily inaccurate. The visual system produces representations of color and shape relatively independently. It is a substantive question what are the constraints on how the visual system brings these feature representations back together. Now, it seems that in fact the psychological constraints on feature binding do recapitulate the metaphysical constraints on the co-instantiation of features. However, the psychological constraints on feature binding cannot be explained just by reference to the nature of the features represented.

One might suggest that the psychological constraint, that one can only feature-bind representations as of spatially coincident features, is not merely a semantic constraint, but that it is still not a structural constraint. To pursue this line, one would have to argue that feature-bound representations as of spatially disjoint properties are *structurally* possible while explaining why they nevertheless cannot be generated by computations in the visual system. On such an account, there would be no structural reason a representation of rectangularity as over there could not in principle be feature-bound with a representation of brownness as over here to form a representation of the rectangularity over there as co-instantiated with the brownness over here; there would be other, non-structural reasons that we never compute such representations.¹⁷

¹⁷Compare with Carruthers (2004, p. 217)'s defense of the claim that honeybees *can* combine certain concepts into beliefs, even though they never do: "From the fact that bees never form beliefs of a certain kind, it doesn't follow that they *cannot*... [I]t might just be that bees are only ever *interested in* the relationships amongst potential new nest sites and the existing colony, and not between the nest sites themselves" (for further discussion of the honeybee case, see Camp, 2009).

Note that the central issue here is not whether there is a computational instead of structural constraint that prevents us from forming feature-bound representations of spatially disjoint features. Structural constraints on representations must at some level be realized by computational constraints on how those representations are generated and used. We can take it as a given that the architecture of the visual system is set up so that it does not compute feature-bound representations as of spatially disjoint yet co-instantiated properties. The central issue is whether the computational constraints that preclude combining representations of spatially disjoint features realize a structural constraint or whether they merely reflect some external, non-structural aspects of visual processing. There are good reasons for thinking that the constraint on feature binding is a constraint on the very structure of feature-bound representations.

I suggest that if the spatial coincidence constraint on which representations get feature-bound is intrinsic to the computations that are primarily dedicated to producing feature-bound representations, then there is good reason to call the constraint a structural one.¹⁸ If the con-

¹⁸We must be very careful about the scope of this consideration. I am not suggesting that any constraint on what representations are formed is a structural constraint so long as that constraint is intrinsic to the computations that are primarily dedicated to forming that kind of representation. The constraint must also, at least, be a constraint on what *constituent* representations can or cannot be combined into a complex representation.

Suppose one computes a representation of a surface's linear size from information about the size of the surface's retinal projection, together with representations of the distance to that surface from oneself and the surface's orientation relative to oneself. Suppose also, for the sake of argument, that representations of linear size are primitive—that is, they do not have constituents and, *a fortiori*, they do not have as constituents representations of distance or orientation. Then, computational constraints on which size representations can be generated given a retinal projection and given distance and orientation representations would not be candidate *structural* constraints. The computational constraints that implement structural constraints must specifically govern the relations between the constituents of the computed representation. Feature-bound representations very plausibly do have feature representations as constituents, and there must be some computational constraints that govern which combinations of feature representations can and cannot be formed. My suggestion is that in this kind of case, the computational constraints likely realize a *structural constraint* if they govern which feature representations can be combined and if the constraints are intrinsic to the computations whose primary function is to form feature-bound representations.

straint is merely a product of what sensory stimuli we receive, of more general-purpose computational mechanisms, or of resource limitations that figure into feature binding, then it is reasonable to suppose that the spatial coincidence constraint is not a structural constraint. (I intend to parallel a standard distinction in generative linguistics: the constraint on feature binding may reflect the central, underlying *competence* of feature binding, or it may reflect features of more general-purpose, peripheral systems.)

It is not plausible that the spatial coincidence constraint is merely a result of what proximal inputs are available to perception. In the first place, aspects of the retinal stimulus alone do not explain why the visual system does not feature-bind representations of spatially disjoint features. If, for instance, a shape and a color project to different regions on the retina, this does not in itself preclude a system from combining representations of those features. Moreover, it is not at all clear what sorts of counterfactual proximal stimuli would cause one to feature-bind representations of spatially disjoint features. And there is no evidence that the spatial coincidence constraint can be overridden by inducing a perceptual set—by altering the subjects' expectations, preferences, and so on. So, there is no reason to think limitations on the inputs to perception are in themselves responsible for the spatial coincidence constraint on feature binding. The constraint must derive at least in part from how one processes retinal stimuli, rather than deriving just from the nature of the inputs to such processing.

Furthermore, it is doubtful that the spatial coincidence constraint on feature binding could be a result of limitations on more general perceptual mechanisms, such as attention or memory, rather than a structural constraint specific to the combination of feature representations. Many other perceptual capacities are not subject to the spatial coincidence constraint. One can

perceive a car in the distance as larger than the pencil in one's hand. This involves combining a representation of the car with a representation of the pencil. Such a combination clearly does not require that the car and the pencil be represented as in the same place, even though representations of relative size may sometimes make use of the same attentional and memory mechanisms into which feature binding taps. I do not see any reason to think that the constraint that feature-bound representations must be of spatially coincident features is inherited from constraints on more general, peripheral systems such as attention. One has to take account of the fact that other perceptual capacities that can make use of those same systems are not subject to the same constraint. The spatial coincidence constraint seems to be intrinsic to the capacity to combine feature representations.¹⁹

I conclude that it is a structural constraint on feature binding that the combined representations must represent spatially coinciding property instances. Representations of spatially disjoint property instances cannot be combined into feature-bound representations. It follows from this constraint that feature-bound representations, abstracting away from their specific constituents, structurally entail the spatial coincidence of the property instances that they represent. Feature-bound representations are committed, by virtue of their very form, to the spatial coincidence of the represented property instances.

¹⁹One might have been tempted to explain the spatial coincidence constraint on feature binding in the following way, as merely a result of basic limitations on attention: feature binding requires spatial attention to the represented features, and spatial attention can be directed only to a single location (or region) at a time. This proposal cannot account for why the spatial coincidence constraint governs feature binding but not other perceptual capacities in which attention may sometimes be involved, such as the perception of relative size.

A further problem with this proposal is that neither of its major assumptions appears to be empirically correct. First, there is a wealth of evidence that feature binding can occur without attention (see Footnote 14). Yet the spatial coincidence constraint continues to operate in preattentive feature binding (Holcombe and Cavanagh, 2001). Second, there is mounting evidence for split spatial attention with spatially discontinuous foci (Pylyshyn and Storm, 1988; Awh and Pashler, 2000; Müller et al., 2003; Scharlau, 2004).

If feature binding structurally entails spatial coincidence, then feature binding cannot be reduced to more familiar modes of combination like conjunction or predicate conjunction. Conjunction and predicate conjunction—indeed, any operation familiar from classical logic—do not structurally entail spatial coincidence or even spatial location. It may be that the sentence, “that is brown and rectangular,” *metaphysically* entails that the brownness and rectangularity spatially coincide, if it is metaphysically necessary that co-instantiated colors and shapes must also spatially coincide to some extent. However, this sentence does not *structurally* entail spatial coincidence. Another sentence of the very same form, “the number two is even and prime,” does not entail that evenness and primeness spatially coincide in the number two, or even that the instances of those properties have any spatial location. By contrast, the implication of spatial coincidence is built into the structural constraints on perceptual feature binding. Conjunction and predicate conjunction lack the structural constraints and structural entailments that characterize feature binding.

The spatial coincidence constraint on feature binding reflects an important regularity in our environment: the sorts of features that we can perceive, such as shape, size, and color, can only be co-instantiated in the same individual if they are located in the same place. The metaphysical regularity, that features such as shape and color can be co-instantiated in an individual only if they spatially coincide, does not in itself preclude perceivers from representing violations of the regularity. However, a perceptual system that could feature-bind representations of spatially disjoint property instances would be capable of making very fundamental kinds of errors, producing representations that could not be accurate in any realistic environment. Feature binding is more reliably accurate because such representations are structurally

disbarred. Effectively, a feature-bound representation's structural entailments guarantee that the way the representation represents this particular scene is a way that an actual scene really could be (given that it is a scene from the sort of environment in which our perceptual capacities evolved). It is good for the perceiver when constraints on the combinability of their perceptual states reflect important environmental regularities. The structural form of feature-bound states reflects a function of those states, which is to represent certain aspects of our normal, physical environment.

3.4 Configural Binding

I now turn to another type of visual structure, *configural binding*, which plays a central role throughout vision. I will argue that this type of structure carries content about spatial patterns whose environmental significance is in fact quite contingent. The discussion of configural binding will illustrate how the arguments given above, that the forms of feature-bound states carry specific spatial content, can be applied to other cases. Different types of perceptual structure, including feature binding and configural binding, will carry their own respective kinds of spatial content. The case of configural binding should also illustrate the broader claim that the structural forms of perceptual states are specially fitted to regularities—in this case, as with feature binding, spatial regularities—in our local environment. In the next section, I will argue that feature binding, configural binding, and any other types of spatially significant perceptual structure all depend on on a more basic, pervasive array-like structure.

In any given scene one perceives discrete individuals. However, one also perceives these

individuals as clustering into certain groups and arrangements. When four books are piled on top of one another, one sees not just four distinct books, but also a stack of books. When one sees a few deer together in the yard, one sees not just the individual deer but also the group that they form. As the Gestalt psychologist Max Wertheimer wrote,

When we are presented with a number of stimuli we do not as a rule experience ‘a number’ of individual things, this one and that and that. Instead larger wholes separated from and related to one another are given in experience; their arrangement and division are concrete and definite. (Wertheimer, 1938b)

The ability to perceive elements as forming groups—the ability of *perceptual grouping*—is a pervasive feature of a normally functioning human’s perceptual life. The early Gestalt psychologists extensively studied the principles governing perceptual grouping. Some of these principles are illustrated in Figure 3.5. The *proximity principle* holds that, all things equal, elements that are near each other will be perceived as members of a common, cohesive group. The principles of *similarity* hold that, all things equal, elements that are similar with respect to color, size, or orientation will be perceived as members of a group. These principles have been refined, made more precise, and supplemented with additional grouping principles (Pomerantz and Kubovy, 1986; Wagemans et al., 2012a,b). It is not entirely clear how the principles interact with each other, but there is some reason to think that they are additive, and that the proximity principle is particularly influential in determining what one perceives as forming a group (Wagemans et al., 2012a).

The grouping principles seem to be at play in many perceptual capacities where the task is

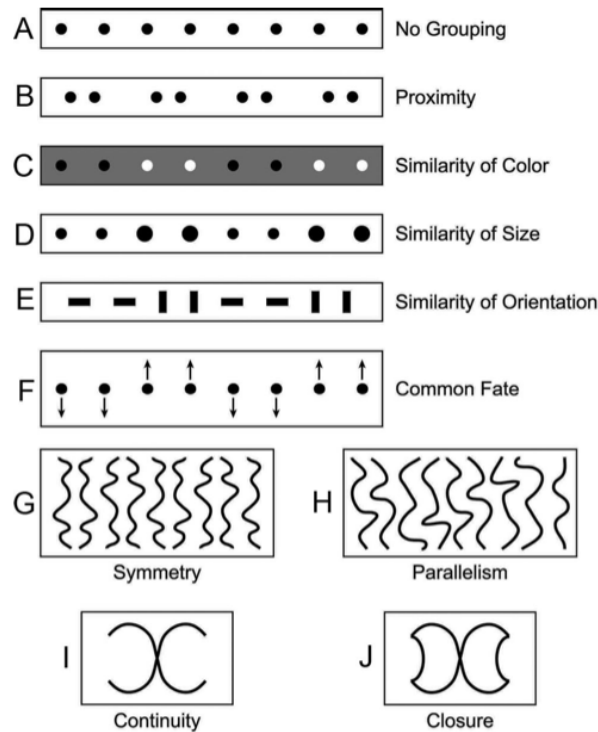


Figure 3.5: Illustrations of perceptual grouping principles. (From Palmer, 1999, p. 258).

to perceive how fragments of a scene form cohesive units. Besides the capacity to perceptually group objects into configurations of objects, grouping principles likely govern our capacity to represent the unified contours of surfaces (Geisler and Super, 2000; Elder, 2015), our ability to perceive individual bodies (Feldman, 1999b), and the ability to perceive the colors of surfaces (Adelson, 1993; Anderson and Khang, 2010), just to name a few.²⁰

I will argue that the principles of perceptual grouping as structural constraints on which representations of individual elements can combine into a unified representation as of a group.

²⁰It is an open question whether there is one unitary mode of combination, *configural binding*, that is at play in each of these capacities, or instead many different modes of combination that are proprietary to the different capacities. It may even be that each grouping principle corresponds to its own mode of combination—so that one has *proximity binding*, *similarity binding*, and so on. For simplicity, I will continue to discuss *configural binding* as a unitary mode of combination. My central claims will carry over, with appropriate modification, to any of these other theoretical alternatives.

This is, in fact, how some contemporary perceptual psychologists have characterized perceptual grouping (see, for example, Palmer, 1977; van der Helm and Leeuwenberg, 1996; Geisler and Super, 2000). If perceptual grouping principles characterize the structural constraints on how perceptual states can combine into representations as of groups, then it follows that perceptual representations as of groups of elements structurally entail that those elements are parts of a cohesive group and that those elements, at the very least, satisfy a disjunction of the grouping principles.

Palmer (1977) conducted a series of experiments supporting the view that perceptual grouping corresponds to a certain mode of combining perceptual states, which I am calling *configural binding*.²¹ Palmer used a synchronic matching paradigm, which consists of the following sort of set-up: a subject is presented with two stimuli and is asked to indicate whether they are identical.²² Palmer used a variation of this paradigm: a *part-matching* experiment in which the subject's task was to determine whether one stimulus was identical to a part of the other.

Palmer's target stimuli were "figures" that consisted of six line segments from a grid (see Figure 3.6). He presented subjects with a target figure alongside a "part probe," consisting of three line segments that may or may not be contained in the target figure. Subjects were asked whether all the line segments in the part probe were present in the target figure or not. The key measure of performance was subjects' reaction times.

²¹I have described these experiments, for a different purpose, in Chapter 1.

²²Diachronic matching tasks, including primed matching tasks, present first one stimulus, then remove it and present either the same or different stimulus, asking the subject whether the most recent stimulus is identical with the previously displayed one.

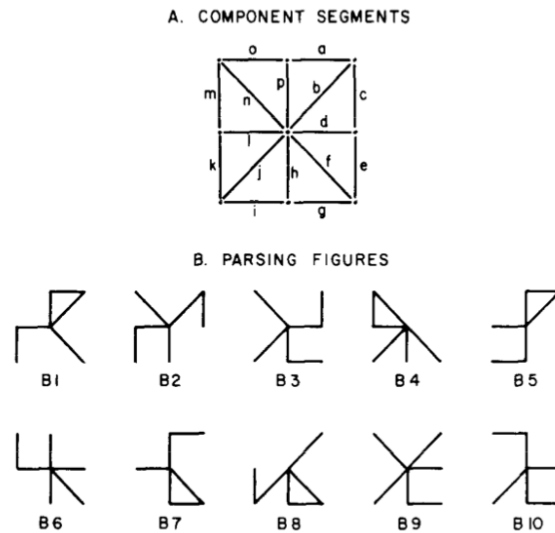


Figure 3.6: Examples of stimuli from (Palmer, 1977)’s study of perceptual organization.

Psychologists commonly offer results from matching studies as evidence that a given type of perceptual state is structured in a certain way from certain kinds of constituent perceptual states (for example Beller, 1971; Duncan, 1984; Kimchi, 2000; Leek et al., 2005). In the case of part matching experiments, the idea is that if subjects are relatively quick to verify the presence or absence of a part probe in the whole figure, it is because that kind of part is represented by a unified representation. The visual system is faster at processing unified, complex representations than at processing a set of uncombined representations. If subjects are comparatively slower to verify the presence or absence of a part probe, then, all other things equal, it is because their representations of the elements of that probe do not combine to form a unified representation, and so their visual system is slower to process the set of representations of those elements. Importantly, matching experiments provide evidence about about the structure of one’s mental states and not just about how one perceive the world to be organized. Reaction times on these experiments reflect the time it takes to process mental

representations, which is assumed to be a function of how those representations are (or are not) structured.

Palmer ranked the different part probes in his experiments by how strongly the line segments seemed to form a perceptual group (what he called the “goodness” of the probe). He measured how strongly the line segments are perceptually grouped by asking subjects, in a separate experiment, how well the line segments seemed to belong together. Supporting the standard claims of Gestalt psychologists, Palmer showed that the reported grouping strength of a set of line segments corresponded to a weighted sum of factors such as how proximal, closed, connected, and continuous the line segments are.

Palmer found that subjects were consistently and categorically faster at detecting the presence of parts in a figure when those parts satisfied a number of the grouping principles. This suggests that the collections of line segments that satisfy a sufficient number the grouping principles (such as the configuration labeled “H” in Figure 3.7) just are the collections that receive unified perceptual representations which can be processed as a unit. By contrast, when a collection of line segments does not, in the context of the whole figure, sufficiently satisfy any of the grouping principles (such as the configuration labeled “L” in Figure 3.7), one’s representations of the individual line segments do not combine with each other to form a unified representation.

The principles characterizing which line segments are perceived as forming cohesive groups, then, also describe how representations of line segments can or cannot combine with each other. A set of representations of line segments can be configurally bound into a rep-

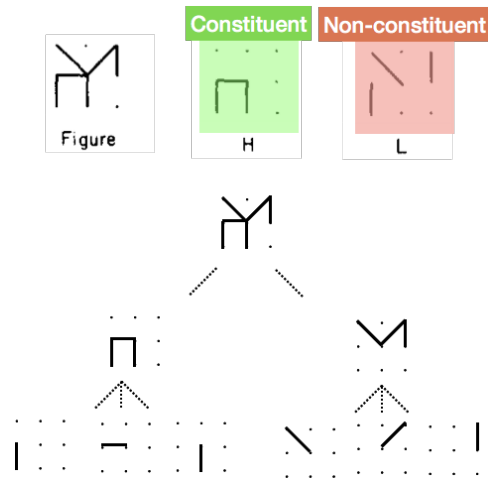


Figure 3.7: Top: A figure from Palmer’s stimulus set (left), along with a set of line segments that receive a unified perceptual representation as a part (middle), and a set of line segments that do not receive a unified perceptual representation as a part (right). Bottom: An illustration of how our representation of the whole figure decomposes into representations of parts of the figure, according to Palmer. Adapted from (Palmer, 1977).

representation of a part only if those line segments, as represented, are sufficiently proximal, closed, connected, continuous, and so on. So, configurally bound representations of groups of line segments structurally entail not only that the represented line segments are members of a group, but also that they are sufficiently proximal, closed, connected, continuous, and so on.

An alternative hypothesis is that representations of groups of elements really have the subject-predicate form *Group*(a_1, \dots, a_n), where a_1, \dots, a_n are representations of particular elements in the group and where the predicate *Group* is only satisfied by individuals that are sufficiently proximal, and so on. On this proposal, it is the content of a particular constituent—namely, the putative predicate or attributive, *Group*—and not the mode of combination of the configural representation that entails that the represented elements be sufficiently proximal and so on. In fact, a number of psychologists describe the representations

involved in perceptual grouping as having merely a logical, subject-predicate form (for example, Watt, 1988; Pylyshyn, 1989; Feldman, 1999b,a).

However, while this proposal may get the accuracy conditions of configural representations right, it does not by itself account for why we are categorically slower to match sets of elements that are not proximal, closed, and so on. Palmer hypothesized that the representations of such elements cannot be combined and so cannot be processed as a unit. However, the current alternative proposes that the representations of such elements can in principle be combined, even if the elements are accurately perceived as far apart, disconnected, dissimilar, and so on. If the representations of these elements can be combined, then the visual system should be able to process those representations as a unit, in which case it should be possible for subjects to respond to such a set of elements just as fast as they do to elements that satisfy the grouping principles.

One could amend the account by arguing that a representation $Group(a_1, \dots, a_n)$, where a_1, \dots, a_n are not as of elements that are, as represented, proximal, closed, and so on, is structurally possible, but that the visual system is, for other reasons, biased against forming such representations. Perhaps one can represent a set of elements as a group even when those elements do not satisfy the grouping principles, if only one were exposed to the right stimuli, or one became sufficiently familiar with such a set, or one learned that the set had functional significance.

As the Gestalt psychologists emphasized, the rules of perceptual grouping are not imposed by the proximal inputs to perception. It is unclear how any aspect of the sensory

stimulus itself, considered apart from the processes operating on it, could explain why one configurally binds some representations but not others.

Perhaps it is not the sensory stimulus itself that constrains configural binding, but rather the familiarity and functional significance of the set of items being represented. There is evidence that familiarity and training can affect grouping. For example, Kimchi and Hadad (2002) found that subjects were able to perceptually group disconnected components of letters in their native alphabet, but could not group disconnected components of nonsense arrangements. Further, well-practiced chess players are able to group configurations of chess pieces that novice players are not able to group (Chase and Simon, 1973). Moreover, the capacity for perceptual grouping develops substantially through childhood and into adolescence (Kovács, 2000; Kimchi et al., 2005).

However, the fact that configural binding develops and can be modulated through perceptual learning is consistent with the central claim that configural binding is subject to spatially meaningful structural constraints. Perceptual modes of combination may develop or be modified over time—they need not be innate and fixed. The relevant grouping principles that, I suggest, govern configural binding may vary to some extent across individuals and within the same individual over time. I propose that at any given time the mode of combination involved in perceptual grouping will have substantive structural entailments, although those structural entailments need not be the same throughout an individual's lifespan. In fact, there seem to be limits on the extent to which training can modulate the standard perceptual grouping principles. Goldstone (2000) showed that subjects could not perceptually group completely random sets of contours, even with many repeated exposures.

Familiarity and functional significance, it appears, can aid in developing and adjusting the parameters of the grouping constraints through perceptual learning. But at any given time, there will be constraints on one's ability to combine representations of elements into representations of groups and these constraints will have substantive semantic import. No stimuli can, even with extensive training, cause one to configurally bind accurate representations of elements that are randomly scattered in the scene.

As with feature binding, one could suggest that constraints on attention, for instance, explain why one cannot configurally bind representations of elements that do not satisfy any of the grouping principles. Perhaps the grouping principles govern how one allocates attention in general, and attention is necessary for grouping. However, attention does not seem to be necessary for perceptual grouping (Rensink and Enns, 1995; Moore and Egeth, 1997; Kimchi, 2009). Moreover, such an account would owe an explanation of how the grouping principles govern the allocation of attention. Just about every extant theory posits that grouping affects attention insofar as attention is allocated on the basis of prior representations of groups in the scene (Baylis and Driver, 1993; Vecera, 1994; Kimchi, 2009). So, it seems that it is in fact constraints on figural binding that best explain patterns of attention, and not *vice versa*.

Palmer's experiments suggest that perceptual grouping principles describe structural constraints on how representations of individual elements in a scene can combine into representations of groups of elements. Substantive environmental principles govern configural binding. One cannot configurally bind representations of just any arbitrarily scattered elements. There do not seem to be any plausible, non-structural explanations. If this is right, then any perceptual representation as of a group will structurally entail that the individuals in the group satisfy

a disjunction (at least) of the grouping principle. Configural binding cannot be reduced to the mere predication of *grouphood*, which would not be subject to the same structural constraints and which would not have such structural entailments.

As was the case with feature binding, the structural constraints on configural binding reflect important regularities in the environment in which our perceptual capacities evolved. Studies on natural scene statistics show that parts of common objects—for example, different parts of a common contour—are very likely to satisfy the perceptual grouping principles, and are far more likely to satisfy those principles than random, unrelated collections of elements in a scene (Brunswik and Kamiya, 1953; Geisler et al., 2001; Elder and Goldberg, 2002). That the parts of the same object tend to satisfy the grouping principles more than disconnected parts of different objects may be a limiting case of a more general regularity: functionally significant collections of elements (objects, groups of prey, and so on) are more likely to satisfy the grouping principles than functionally insignificant collections of elements (see Feldman, 1999b, 2003). The structural entailments of configurally bound representations provide reliable assurance that, supposing the constituent representations are accurate, the elements that one represents as forming a group are parts of a functionally meaningful configuration.

The environmental regularities here are far more contingent than in the case of feature binding. Whereas it seemed highly probable, and perhaps even metaphysically necessary, that anytime a shape and color are co-instantiated in the same individual, they must be located in the same place, it is quite contingent that in our environment it is the close together, similar things that are most significant to one's survival. Configurally bound states have substantive structural entailments that are highly specialized for accurately representing the environment

in which our species evolved.

3.5 Visual Arrays, Revisited

The arguments in the previous two sections can be marshaled to support the view that array-like representations—representation's that *structurally* entail the locations (or at least directions) in which perceived elements are located, as well as topological relations between those locations—play a central role in vision. By positing that feature binding and configural binding operate over array-like representations, one can produce a unified explanatory framework for understanding these different perceptual modes of combination.

Consider the constraints on feature binding and configural binding. These dictate that representations of property instances can be feature-bound only if the represented locations of those property instances coincide, or that representations of elements can be configurally bound only if the represented locations of those elements are sufficiently proximal, and so on. Strictly speaking, these constraints are not appropriately formulated. As stated, the constraints refer to the contents of the to-be-combined perceptual states—how they represent the spatial positions of property instances, for example. But the constraints must apply in virtue of *structural features* of the to-be-combined perceptual states. It is the structural features of a representation that explain how it can and cannot combine with other representations.²³

A natural way to reformulate the constraints on combination in terms of structural features

²³I do not assume that reference to a state's content and reference to its structural properties are mutually exclusive. Structural properties may correspond with semantic properties. Some have even argued that a mental state's structural properties metaphysically depend on its semantic properties (Burge, 1986; Peacocke, 1994; Rescorla, 2012).

of representations is to describe those representations as occupying cells in a representational array, where those cells correspond to the lines of sight (and perhaps distances) in which denoted items are represented as being located. One can then define the rules for admissible combinations of perceptual states in terms of the relations among the cells or addresses that those representations occupy. For example, in order to feature-bind two representations, the representations must occupy the same cells in the array (or enough of the same cells, or adjacent cells, as the case may be). And, all things equal, in order to configurally bind a set of representations, those representations must occupy nearby cells in the array. The picture one winds up with is one of an underlying set of representations that are structurally related within the array, which can then be combined with one another to form more complex feature-bound and configurally bound representations (see Figure 3.8).

In fact, theories of feature binding and configural binding standardly posit what look, on the face of it, to be array-like representations. Most theories of feature binding suggest that *feature maps*, in which representations of features are indexed according to the perceived locations of those features, are central to the mechanism underlying feature binding (Treisman, 1988; Quinlan, 2003). Similarly, many theories of perceptual grouping assume that items are in some way indexed or addressed by their locations, and that constraints on combinations of representations are sensitive to these indexes or addresses (Palmer, 1983; Vecera, 1994; Geisler and Super, 2000).

However, it can often be ambiguous in these theories whether the postulation of a “feature map” or of a representational coordinate system serves merely to characterize the spatial contents of the representation in question, as Pylyshyn might claim, or whether it also

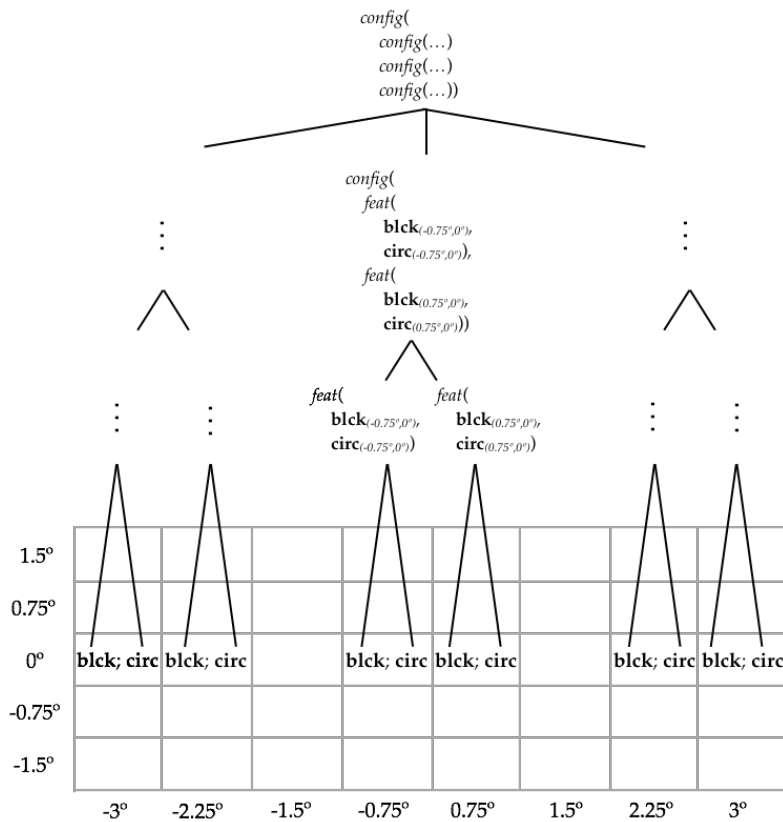


Figure 3.8: An illustration of how feature binding and configural binding operate within a basic array-like structure. Primitive feature representations (“**black**,” “**circle**”) occupy cells in the array, addressed by the line of sight (given in degrees of visual angle). Representations that occupy the same cell can be feature-bound (indicated here by a structural operator, *feat*), and representations in nearby cells can be configurally bound (indicated by a structural operator, *config*). Configurally bound representations that cover nearby sets of cells can themselves be configurally bind.

characterizes the structural form of the representation. Some of the confusion over which interpretation is appropriate stems from confusion between the structural and causal roles of spatial representations. Theoretical discussions of feature binding and configural binding often characterize the role of spatial representations by positing some axioms or rules that govern transitions from representations of features at locations to representations of those features as co-instantiated in an individual, or from representations of individual items at locations to

representations of cohesive groups (see, for example Treisman, 1988; Tacca, 2011; Feldman, 1999b,a). However, it is important to distinguish between rules that govern the transitions from one representation to another and the rules that govern how representations can and cannot combine (see Chapter 1).

Part of the force behind Pylyshyn's challenge stems from the fact that people have often attended primarily to the role that represented spatial relations plays in explaining constraints over the transitions between representations. Those processing constraints do not immediately imply that the represented spatial relations are embodied in the structural forms (that is, modes of combination) of the representations being processed, rather than being represented merely by the constituent parts of those representations.

In the previous two sections, I employed a common strategy to identify which aspects of a representation's content are contributed by the *form* of that representation, abstracting from its parts. If the relevant semantic properties are associated with the constraints on how representations can combine, then those semantic properties are implied by the relevant mode of combination (and so by the relevant structural form). I argued that the constraints on feature binding and configural binding are not well-explained by peripheral constraints on processing—constraints imposed by the nature of the stimulus or by more general systems such as memory and attention. All things being equal, if the constraints governing which combinations of representations are possible are intrinsic to the mechanisms dedicated to forming those combinations, then those constraints reflect genuine structural constraints. I argued that the rules governing feature binding and configural binding each make reference to represented spatial relations. Spatial coincidence, in the case of feature binding, and relations such as

proximity, in the case configural binding, are implied by the structural constraints on which representations can and cannot be feature bound or configurally bound. Since the spatial relations figure into structural constraints on representations, we are warranted in saying that they are entailed by the very structural forms of those representations.

By focusing in particular on *structural constraints* on combination, and not merely on *processing constraints*, one can better differentiate which aspects of a representation's content are contributed by its form. The best statement of the structural constraints on feature binding and configural binding will posit a basic array-like structure. The combinatorial constraints governing feature binding and configural binding apply to visual states in virtue of how those states occupy cells in the representational array. A good reason to think that visual states have array-like structure is that this offers a unified scheme into which other modes of combination can naturally fit.

3.6 Form and Function

I expect that the same sorts of empirical arguments I have given for the claims that feature-bound and configurally bound representations have spatial structural entailments can be given for the kinds of visual representational structure involved in contour completion (Kellman and Shipley, 1991), part-based representations of shapes (Hoffman and Richards, 1984; Singh and Hoffman, 2001), and framework-anchoring in lightness and color perception (Gilchrist et al., 1999; Gilchrist, 2006). Similar arguments can also be made for feature binding and configural binding as it occurs in auditory perception (Denha and Winkler, 2015) and touch

perception (Kappers and Bergmann Tiest, 2015). I expect that these different capacities all involve types of structures that carry content about relevant spatial relationships. One can provide a unified framework for understanding how perceptual states can combine into these different structures by supposing a basic array-like structure over which these other modes of combination are defined.

I now want to offer some reasons to expect that these arguments can be generalized in a fairly robust way across modalities and across animals. In the cases discussed above, constraints on structural combination corresponded to natural regularities in our local physical environment that were relevant to the perceptual capacity in question. In the case of feature binding, the relevant environmental regularity was that perceptually discriminable, co-instantiated features must spatially coincide. In the case of configural binding, the relevant regularity was that elements of functionally significant groups satisfy a disjunction of the perceptual grouping principles. In effect, all the types of perceptual representations discussed here structurally entail that relevant environmental regularities are satisfied by the items being represented.²⁴

The structural entailments of feature binding and configural binding facilitate the functions of accurately representing co-instantiated features and functionally significant groups of objects. Perceptual capacities function to represent important aspects of one's immediate

²⁴Shimojima (1996, 2001) suggests that it is distinctive of *iconic* representations that law-like constraints over those representations correspond to law-like constraints in the represented subject matter. On this criterion, perceptual representations are iconic. The iconicity of perceptual representations is usually associated with their purported array-like structure. However, the constraints governing feature binding and configural binding taken on their own, correspond to constraints in the environment. So feature binding and configural binding should also count as iconic according to Shimojima's definition. If one adopts Shimojima's position, then the iconicity of perception is not limited to the presence of array-like representations.

physical environment—different features that belong to a particular individual, meaningful groups of objects, and so on. There are statistical regularities in how co-instantiated features and meaningful groups of things are spatially arranged. The very structures of our perceptual states help ensure that we are representing things that satisfy those statistical regularities.

Perceptual capacities in different modalities and in different animals rely on their own relevant environmental regularities and patterns in more accurately representing a particular aspect of the immediate environment. These regularities may be spatial, temporal, or otherwise. It will regularly be the case that it is beneficial for perceptual states to structurally reflect the relevant regularities. In general, a creature does not have to worry about representing an impossible, improbable, or evolutionarily useless situation, if representations of such situations are not even structurally possible. By structurally entailing that relevant regularities obtain, perceptual states are guaranteed a baseline of accuracy and functional relevance. Even when perception gets it all wrong about a specific scene, one at least does not represent an impossible, highly improbable, or functionally meaningless scene.

I have argued that the constraints on feature binding and configural binding cannot be explained by reference to non-structural factors, such as the nature of the relevant sensory stimuli or limitations on general systems, including attention and memory. I can only speculate that this will generally be true. Recall, first, that nothing in the sensory stimulus *requires* one to only feature-bind representations as of spatially coincident elements, or to only configurally bind representations of nearby (or similar, etc.) items. In general, sensory stimuli vastly underdetermine what the perceptual system can do with that stimuli. I think it is unlikely that constraints on combination will normally be due to limitations inherent in the stimuli

themselves.

The main question is whether such constraints might be inherited from limitations on general-purpose systems such as attention and memory. I think this is unlikely to be the norm. Systems such as attention and memory typically function to subserve many different perceptual and psychological capacities. These different capacities may each benefit from structurally embodying regularities that are relevant to those specific capacities. If constraints on the operation of a general-purpose mechanism reflect certain environmental regularities, then any capacities that make use of that mechanism will also reflect those regularities, even if the regularities are irrelevant or overly specific for the functioning of those other capacities. It may well be more beneficial in general for the more special-purpose capacities, such as the capacities for feature binding and configural binding, to embody the relevant regularities through structural rules than it is for more general-purpose systems to embody those regularities through mere processing constraints.

There some reason to expect that for most perceptual capacities in most animals, perceptual states will structurally entail ecologically significant types of arrangements—that is, arrangements the tracking of which is important for the creature’s survival, or for its accurately representing important features of the environment. The structural forms of perceptual states, in general, are set up to facilitate the specific functions of those perceptual states—to represent particular features of the creature’s local physical environment.

3.7 Forms of Mind

One can find comparisons, in both philosophy and psychology, between perceptual representations and formulas of classical logic. Rock (1985), Watt (1988), Pylyshyn (1989), and Feldman (1999b,a) all characterize perceptual representations as subject-predicate representations of the form one finds in classical predicate logic. And philosophers regularly debate over what type of expression in classical predicate logic best captures the form of perceptual content—for example, whether perceptual contents are best captured by existentially quantified expressions or open sentences (Schellenberg, 2018). These comparisons can yield insights into the structure and semantics of perception. However, if the arguments in the previous sections are correct, then any attempt to characterize the semantics of perceptual states in terms of formulas of classical logic will be fundamentally inadequate. The forms of classical logic are not subject to the same structural constraints and do not have the same structural entailments as those of perception. Perceptual states may, like subject-predicate expressions in logic, refer to individuals, indicate properties and relations, and attribute such properties and relations to those individuals (see Burge, 2010a). However, the structural forms of perceptual states standardly build in spatial (and perhaps other) content that is lacking in the more abstract form of mere predication.

It is perfectly natural that the forms of classical logic should not have the sorts of structural entailments that characterize perception. One of the hallmarks of classical logic is its generality. The very same logical form can be used to represent many different topics. The subject-predicate form, $\lceil F(a) \rceil$ can be used to represent that spiders are scary, that the number

two is prime, and that politics is depressing. I have argued that the structural forms of perception are far more specialized, reflecting the specialized functions of our perceptual capacities. The structural entailments of perceptual representations commit those representations to being about specific kinds of spatial (or other) arrangements (cf. Cummins, 2010; Camp, 2007). Perceptual capacities trade off the general applicability of logical forms in order to more reliably represent the specific things—co-instantiated features, functionally significant groups, and so on—that those capacities function to represent.

The structural forms of perception cannot be reduced to the forms of classical logic. It may be possible to replicate the accuracy conditions and state transitions of perception by defining a classical first-order language with suitable predicates and axioms. However, while this language may get the accuracy conditions and patterns of transitions right, it will get the structural rules wrong and it will misallocate semantic significance to constituent terms that properly belong to modes of combination.

Examining the semantic significance of structural forms in perception does not just help us better understand perception itself, but can give us greater purchase on a central problem in philosophy of mind: how does perception differ from thought? Whereas perceptual states are structurally specialized to certain types of ecologically important patterns, there is good reason to think that thoughts have more general, abstract structural entailments. We are capable of thinking about and relating wildly different kinds of subject matters. In science, we develop unified theories of phenomena that are superficially quite different (for example, an apple's falling to the ground and the moon's orbiting the Earth). Mathematicians explore the properties of numbers without, on the face of it, presupposing that numbers are concrete, spatial

objects. A food critic may compare the taste of a lamb tagine to the sound of a jazz trumpet playing over a bass. Story tellers might imagine that their hero or heroine gets a migraine and a cramped foot after hurdling into a black hole. Thinkers may even entertain contradictions when identifying the inconsistency in a theory, so as to reject it. There must be structural forms available to thought that are abstract enough in their structural entailments so as to allow one to flexibly and creatively integrate content about arbitrarily disparate topics, and even about contradictions (for similar points, see Camp, 2007).

Indeed, a number of people have suggested that thoughts are structured like formulas in classical logic (Fodor, 1975; Rey, 1995; Devitt, 2006). There is some plausibility to this, because human thought seems to have the same general applicability as classical logic is often thought to have. Perception differs importantly from both thought and logic in that the structural forms of perception are far more specialized in their applicability.

A point of clarification is necessary here. My arguments that perceptual representations have spatially significant structure hinge on claims about how certain perceptual representations can and cannot be combined, and I have just suggested that the structural forms of perception are far less general in their applicability than are the forms of thought. However, I am not claiming that perceptual capacities violate what Evans (1982) called *the generality constraint*. With the generality constraint, Evans intended to provide a necessary condition on a representational capacity's being a conceptual capacity. Evans suggested that conceptual capacities (for example, our capacities to think certain thoughts) are structured. If we have a set of structured representational capacities, Evans suggested, then the constituent sub-capacities must be systematically recombinable. For example, if I can think that *Elisabet is happy*, and

if I can think that *Alma is sad*, then I must be able to think that *Elisabet is sad* and that *Alma is happy*.

My arguments have not assumed that any perceptual capacities violate the generality constraint, at least as it has standardly been interpreted. While I appeal to constraints on how perceptual representations (of features, or of individual elements in a scene) can be combined, I do *not* claim that those representations fail to be systematically recombining. Rather, my arguments have pointed to constraints on how perceptual states *can* be systematically recombined. Even Evans accepted that there must be some structural constraints on recombining. Consider Max Ernst's sentence, "Price they are yesterday agreeing afterwards paintings" (quoted in Camp, 2004). Each word in the sentence expresses a concept, but the string as whole does not express a unified thought (although one could associate a thought with the sentence). The generality constraint requires conceptual capacities to be recombining *within the bounds of structural possibility* (see Clapp and Duhau, 2011). My arguments have rested on identifying the bounds of structural possibility in perception and showing that these boundaries endow perceptual states with spatial commitments about how things are spatially (and perhaps otherwise) arranged.²⁵

Examining the types of semantic significance embedded in the structural forms that characterize our mental capacities offers a useful approach for elucidating underlying differences between those capacities. Perceptual capacities may lie at one end of a spectrum, with quite

²⁵ Additionally, the generality constraint was intended to be a condition on *conceptual* representation, whereas my concern here is not with whether perceptual representations are conceptual (although, ultimately, I do not believe that they are). It seems to be perfectly possible to have conceptual representations that figure into semantically rich structural forms—natural languages may contain examples of such cases.

specific, ecologically tuned structural entailments, while human thought lies at the other end, with far more general, abstract structural entailments.

CHAPTER 4

The Perspectival Character of Perception

Perception . . . is from a perspective. But it does not represent appearances or perspectives.

Tyler Burge

ORIGINS OF OBJECTIVITY

In perception, one can transcend the limitations imposed by one's perspective and see things, in many respects, as they really are. For example, a circular coin can reflect dramatically different patterns of light to your eyes depending on where you are viewing it from. Each of these patterns of light could have been reflected by infinitely many different types of surfaces of different shapes and sizes. On the face of it, these ever-changing and always ambiguous patterns of light at the eye carry little information about the shape and size of the coin. Yet you normally can see the coin as having a unique size and shape. The size and shape you see the coin as having normally remains the same from one viewpoint to the next—this is known as *size constancy* and *shape constancy*. And, normally, you can see the size and shape of the coin accurately. But while you can perceive the size and shape of the coin accurately from just about any perspective, your perception of the coin remains marked by your perspective on it. What accounts for the perspectival character of perception?

Two types of phenomena exemplify the perspectival character of perception. Take a circular coin and rotate it away from you in depth. In the first place, your perception of the coin changes in some respect as you rotate it. Loosely speaking, the coin has a different look when seen head-on than when seen at an angle even though we see it as circular in both cases. Call this sort of phenomenon *perspectival variance*.¹ Second, it is often said that there is some respect in which your perception of the slanted coin is similar to your perception of a head-on ellipse. Loosely speaking, even as the coin looks circular to you, there is some sense in which it also has an elliptical look. Call this sort of phenomenon *perspectival similarity*.²

Perspectival variance and similarity can be found throughout perception.³ The perception of non-geometrical properties such as color and lightness exhibits patterns of perspectival variance and similarity.⁴ For example, your perception of a white surface under shade differs from your perception of that white surface under sunlight (perspectival variance), and some hold that your perception of the white surface under shade is similar in some respect to your perception of a darker gray surface that is under sunlight (perspectival similarity), even if your perceptions of both are fully accurate (see, for example, Chalmers, 2006, p. 87). Analogous cases arise in hearing and touch. Despite the many differences between these forms of per-

¹This sort of phenomenon has also been called “perceptual relativity,” “situation-dependency,” “nonconstancy,” and “inconstancy.”

²While most people acknowledge that there is perspectival variance, some doubt that there is perspectival similarity (Smith, 2002; Hopp, 2013, see, for example). For the purposes of this paper, I will treat perspectival similarity as real. The core points in this paper apply even if there is only perspectival variance.

³Another commonly cited example in spatial visual perception: two same-sized trees at different distances look different and the farther tree looks smaller in some sense—that is, it looks similar to a nearer, smaller tree (see Peacocke, 1983).

⁴The *lightness* of a surface is the degree to which the surface is white (fully light) or black (fully dark). Lightness is a property of the surface, independent of how it is illuminated.

ception, they all raise the same general puzzles. Given that one can accurately perceive the sizes, shapes, and colors of things under many different conditions of observation, why should one's perceptions of these properties vary depending on those conditions? And why should there be any relevant similarity between one's accurate perceptions of two different properties (circularity and ellipticality, for example) when each is presented in a different condition of observation (slanted and head-on, respectively)? If, in perception, one can transcend the ever-changing flux of equivocal sensations and represent the more or less fixed properties of things in the world, then why does perception reflect that changing, equivocal flux, in the form of variance and similarity?

The standard explanations of the perspectival character of perception take the form: *Perception is perspectival because we perceptually represent certain relational properties in addition to non-relational properties such as size, shape, and color.* The relational properties in question are supposed to constitute the varying "looks" or "appearances" that objects can have relative to different conditions of observation. I will reject this sort of explanation and propose that a better explanation takes the form: *Perception is perspectival because perceptual representations have a certain sort of structure.*

I begin in Section 4.1 by reviewing the arguments that the perspectival character of perception depends on the representation of relational properties. I go on to argue against the two standard approaches to characterizing the relevant relational properties. In Section 4.2, I discuss the *pluralist approach*. According to the pluralist approach, while the perspectival character of visual spatial perception may depend on our perceptual representations of one kind of relational property, the perspectival character of color perception may depend on our

perceptual representations of another kind of relational property, and likewise for hearing and touch. I argue that the pluralist approach is unsatisfying, since it does not account for the seemingly unified nature of the perspectival character of perception. In Section 4.3, I introduce the *perspectival properties approach*, which holds that all perspectival forms of perception involve representations of a distinctive sort of property: *perspectival properties*. I argue, in Section 4.4, that this approach violates central commitments of empirical psychology. In Section 4.5, I offer an alternative type of account. Both the pluralist and perspectival properties approaches attempt to explain the perspectival character of perception in terms of what properties our perceptual states represent. I propose, instead, that the perspectival character of perception depends on the way our perceptual states are structured.

4.1 Slanted Coin Arguments

I take for granted that perception is representational. Perceptual states have content about objects in the world and their properties. For example, my current perceptual state represents the tabletop in front of me as brown and rectangular. The content of my perceptual state sets a condition on how the world has to be if that state is to be accurate: my current perceptual state is accurate just in case the tabletop is in fact brown and rectangular.

Perceptual representations exhibit perspectival variance and similarity. A coin has a different look when seen at an angle than when seen head-on in part because one's perceptual representations of the slanted coin and of the head-on coin differ in some respect. The slanted coin has a similar look to a head-on ellipse because one's perceptual representations of the

slanted coin and of the head-on ellipse are similar in some respect.⁵ What features of perceptual representation give rise to variance and similarity?

Let us make this question more precise. We can describe perspectival variance and similarity in terms of the representational features of perceptual states. I will say that a *representational feature* of a perceptual state is a feature that is essential to the way that state represents its subject matter. So my current perceptual state has the representational feature of representing the tabletop. More specifically, this state has the representational feature of representing the tabletop as *brown* and as rectangular; and it has this feature because it has the representational feature of being a combination of a representation as of an instance of *brownness* and a representation as of an instance of *rectangularity*. Perspectival variance consists in the fact that one's perception of a slanted coin (or a shaded white surface) and one's perception of a head-on coin (or an unshaded white surface) will have different representational features. Perspectival similarity consists in the fact that one's perception of a slanted coin (or a shaded white surface) and one's perception of a head-on ellipse (or an unshaded gray surface) will share relevant representational features. An explanation of the perspectival character of perception should specify what kinds of representational features account for perspectival variance and similarity.

The standard accounts of the perspectival character of perception suppose that perceptual

⁵Some have argued that only the phenomenal character of one's perceptual states exhibits variance and similarity (see Peacocke, 1983). *Representationalists*, who hold that the representational content of a mental state fully determines that state's phenomenal character, have responded by offering explanations of how perceptual representation is perspectival as well. In this paper, I take for granted that perceptual representation is perspectival and that perspectival variance and similarity arise from how we represent the world. I will leave it open as to how the perspectival character of perceptual representation relates to the perspectival character of our perceptual phenomenology. I will not comment on the broader question, with which representationalists are concerned, of how the representational content of an experience relates to that experience's phenomenal character.

representations of the head-on coin and the slanted coin have different representational features insofar as they attribute different relational properties to the coin.⁶ Likewise, these views suppose that one's perceptual representations of the slanted coin and the head-on ellipse share representational features insofar as they attribute the same relational properties to the slanted coin and the head-on ellipse. Let us review the arguments for these suppositions.

4.1.1 The Argument from Variance

Consider perspectival variance. A common line of thought goes like this: as you rotate a coin, your perception of the coin changes even as you continue to perceptually represent the coin's circularity, which is a non-relational property of the coin. Since you continue to perceptually represent the coin as circular in shape, your perception must not be changing with respect to your representation of the coin's shape. But notice that as you rotate the coin, some of its relational properties, including its orientation with respect to you, change. The change in your perception must, so this line of thought continues, be due to your representing the coin as having different relational properties at different times.

There are some gaps in this argument that need to be filled in. For example, the point is

⁶It is not a trivial task to explicate the notion of a relational property. For our purposes, we can consider relational properties as properties whose instantiation in an entity depends metaphysically on how that entity is related to other entities. For example, being slanted and being shaded are relational properties whose instantiation in a surface depends, respectively, on how that surface is related to a viewer and how it is related to an illumination source. I take it that there is a fundamental contrast between relational properties such as the orientation and illumination of a surface, and properties such as the shape and color of a surface. Throughout this paper I will describe shape and color as *non-relational* properties—that is, properties whose instantiation in an entity does not depend on how that entity is related to other entities. One may question whether color, or even shape, are truly non-relational. However, little of the substance in this paper hinges on these questions, so long as we have some way of making sense of a basic contrast between properties such as orientation and illumination and properties such as shape and color (for discussion of these issues, see Byrne and Hilbert, 2003; Cohen, 2004; Skow, 2007).

not just that our representation of the coin's circularity remains constant, but that our representation of all of the coin's relevant non-relational properties remain constant. Further, the argument assumes that the representational difference in the perceptions of the head-on and slanted coin is best explained by a difference in what properties are represented. Since you represent the head-on and slanted coin as having the same non-relational properties, the argument goes, the difference in your perceptions is that you represent the coin as having different relational properties.

The argument can be formulated more adequately as follows. Let "slanted coin" denote an accurate perceptual representation of the circular surface of a coin that is slanted in depth relative to one's line of sight under normal conditions of observation. Let "Head-on coin" denote an accurate perceptual representation of the circular surface of the coin were it oriented head-on under normal conditions.

(V1) Slanted coin and Head-on coin have different representational features.

(V2) The best explanation for why Slanted coin and Head-on coin have different representational features is that they do not represent all the same properties.

(V3) Slanted coin and Head-on coin represent all the same relevant non-relational properties (for example, circularity).

(V4) So, the best explanation for why Slanted coin and Head-on coin have different representational features is that there is some relational property that is represented by one but not the other.

Going forward, I will assume V1 and V3 to be true. The difference between one's percepts of the slanted coin and the head-on coin seems to be a difference that bears on how one represents the coin, though it is not a difference in what non-relational properties one represents the coin as having. We can construe V2 as a claim either about what properties are denoted by one's perceptual states or about what properties enter into the perceptual modes of presentation of shape properties (see Chalmers, 2004; Thompson, 2010). V2 is perhaps motivated by the idea that the representational features of an experience are entirely determined by what properties the experience represents. According to that view, the only possible explanation for the representational difference between perceptual states is that they represent different properties (and perhaps different objects). But one might find V2 plausible without adopting such a restrictive account of representational features. As a general methodological point, it is often more straightforward to discover and theorize about what objects and properties are psychologically represented than to discover and theorize about other sorts of representational features. Further, one is often, though perhaps not always, experientially aware only of the objects and properties that one perceptually represents. So, it is reasonable to expect that when one is aware of a difference in one's experiences of a head-on coin and the slanted coin, what one is aware of is a difference in the properties possessed by the head-on and slanted coin (see Hill and Bennett, 2008). I take V2 to be a reasonable but defeasible methodological assumption.

Notice that perspectival variance does not imply perspectival similarity. In principle your perception of the slanted coin could differ from your perception of the head-on coin without being similar in any respect to our perception of the head-on ellipse. As we will see, some

explanations of variance cannot also account for similarity. So we should distinguish between the argument from variance and a parallel argument from similarity.⁷

4.1.2 The Argument from Similarity

The usual formulation of the argument from similarity goes like this: there is some sense in which your perception of the slanted, circular coin, unlike your perception of the coin when it is oriented head-on, is similar to your perception of an appropriately shaped, head-on ellipse—the slanted coin has an “elliptical look,” so to speak. Assuming that you accurately perceive the coin as circular and that your perception is not illusory, you do not also perceive the coin as elliptical. So you do not perceptually attribute the same shape to the slanted coin and head-on ellipse. There must be some relevant relational property that you perceive both the slanted coin and the head-on ellipse as having.

As with the argument from variance, there are gaps in this reasoning. The argument assumes that representational similarity is best explained by a similarity in what properties one represents. The thought is that because one does not perceptually attribute the same shape, or any other relevant non-relational property, to the slanted coin and the head-on ellipse, it must be that one attributes some relational property to both. Let us fill out the argument. Let

⁷Many philosophers freely shift between talking about variance and similarity. They assume that both perspectival variance and similarity are expressions of the same underlying phenomenon. But not everyone shares this assumption; some acknowledge perspectival variance but deny perspectival similarity. For example, Smith (2002, p. 181–2) claims that while one’s experience of a coin’s shape does differ when one views the coin from different angles, there is no relevant similarity between the perception of a slanted coin and that of a head-on ellipse. One cannot appeal to perspectival variance to prove that perspectival similarity is real, as Jonathan Cohen does when he writes in response to Smith: “Like it or not, inconstancy reactions [variance] are part of the observed data; they show that inconstancy [variance] (for example, along the dimensions of size, shape, color) is part of subjects’ phenomenal experience in perceiving the cases of interest” (Cohen, 2010, p. 100).

“Slanted coin” and “Head-on coin” mean what they did in the argument from variance, and let “Head-on ellipse” denote an accurate perceptual representation of an appropriately shaped elliptical figure, viewed head-on under normal conditions.

(S1) There is some representational feature that both Slanted coin and Head-on ellipse have, but which Head-on coin does not have.

(S2) The best explanation for why there is some representational feature that both Slanted coin and Head-on ellipse have is that there is some property that they both represent.

(S3) All the non-relational properties that both Slanted coin and Head-on ellipse represent are also represented by Head-on coin (for example, *having a bounded surface*).

(S4) So, the best explanation for why there is some representational feature that both Slanted coin and Head-on ellipse have, but which Head-on coin does not have, is that there is some relational property that both Slanted coin and Head-on ellipse represent, but which Head-on coin does not represent.

The arguments from variance and similarity each give us reason to think that an explanation of the perspectival character of perception must involve reference to representations of relational properties. In the next sections, I discuss different approaches as to what those relational properties are. After considering weaknesses in these approaches, I will recommend that giving up V2 and S2 and pursuing an alternative type of explanation.

4.2 The Pluralist Approach

Let us begin by considering what I will call the *pluralist approach* to explaining the perspectival character of perception in terms of the representation of relational properties. The pluralist approach applies a divide-and-conquer strategy to explaining why different types of perception are perspectival. A pluralist account might explain the perspectival character of spatial perception in terms of perceptual representations of one kind of relational property, while it might explain the perspectival character of color perception in terms of perceptual representations of another kind of relational property.⁸ Advocates of the pluralist approach tend to prefer explaining the perspectival character of perception in terms of representational capacities that figure into existing psychological theories and that are mentioned in standard textbooks (for example, Palmer, 1999; Frisby and Stone, 2010). I will argue that the pluralist approach's lack of a unified explanation of the perspectival character of perception is unsatisfying.

Consider the slanted coin. What is the representational difference between your perception of the head-on coin and your perception of the slanted coin? One answer that a proponent of the pluralist approach might give is that you represent the head-on coin as *head-on*, whereas you represent the slanted coin as *slanted*. It is well established that we perceive how surfaces are oriented with respect to us. So the pluralist approach seems to have a straightforward and plausible explanation for this case of perspectival variance.

While an appeal to the perception of surface orientation can explain perspectival variance in this case, it cannot explain perspectival similarity. Why is your perception of the slanted

⁸Examples of the pluralist approach can be found in (Tye, 1996, 2005; Matthen, 2010; Hopp, 2013).

coin similar to your perception of a head-on ellipse? Under normal conditions, you perceive the slanted coin as slanted and the head-on ellipse as head-on. The similarity in these perceptions cannot be due to your perceptually attributing similar orientations to the coin and the ellipse.

In order to account for perspectival similarity, the proponent of the pluralist approach might appeal instead to representations of the *egocentric directions* to points on a surface—that is, the directions in which the points on a surface are located with respect to one’s viewpoint. It is well-established that one can perceptually represent points in a scene as located at certain distances away from oneself and in certain directions from one’s viewpoint. Now, consider the directions (but not distances) from your point of view to the points on the slanted coin and to the points on a head-on ellipse that perfectly occludes the slanted coin. The directions to the points on the slanted coin will be the same as the directions to the points on the occluding ellipse. To illustrate this, point your finger to the edge of a slanted coin and let your finger trace the outline of the coin. The path of your finger passes through every direction from your shoulder to the outline of the coin. That path—that set of directions—will be the same for an appropriately shaped head-on ellipse. The proponent of the pluralist approach can claim that your perceptions as of the slanted coin and of the head-on ellipse are similar because they both involve representations as of the same set of directions from your viewpoint to the points on the represented surface.

In fact, this proposal needs a slight modification. Suppose the slanted coin is adjacent to the head-on ellipse. The set of absolute directions from you to the points on the two surfaces will be different because one set of directions will point toward the coin and the other toward

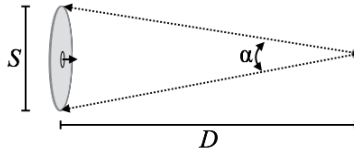


Figure 4.1: The visual angle subtended by two or more points on a surface is the angle α between rays extending from the viewpoint to those points on the surface. The visual angle subtended by points on a surface depends on the distance S between those points, the orientation of the surface relative to the viewpoint, and the distance D between the surface and the viewpoint.

the ellipse. The explanation should appeal instead to what is in common between these sets of directions—namely, the *relations* among the directions. The *visual angle* of a surface corresponds to the relevant relations among the directions from one’s viewpoint to points on the surface of an object. Roughly, the visual angle of a pair of visible points on an object is the angle or difference between the directions from a point of view to those two points (Figure 4.1). The visual angle of a surface as a whole corresponds to the set of visual angles between every pair of visible points on the surface, or to the shape of a cone whose base is the visible portion of the surface and whose apex is the viewpoint.

The head-on coin has a different visual angle than the slanted coin, which in turn has the same visual angle as a head-on ellipse. So, representations of the visual angles of surfaces are well-suited to account for both perspectival variance and similarity in size and shape perception (see Tye, 1996; Jagnow, 2012). Moreover, there is some precedence in the vision science literature for the claim that we perceptually represent visual angles (see McCready, 1985; Kaneko and Uchikawa, 1997). A representation of the visual angle between two (or more) points in a scene could subserve the perception of size and even guide actions by specifying, for example, how much one’s eye or head would have to rotate in order to shift focus from one

point to the other.

Representations of visual angles cannot, however, explain other cases of perspectival variance and similarity. For example, one cannot appeal to the perception of visual angles in order to explain why there is some respect in which one's perception of the shaded side of a uniformly painted wall is different than one's perception of the unshaded side or why there is some respect in which the shaded white surface is similar to one's perception of an unshaded surface with a darker paint. The proponent of the pluralist approach will have to find some other way of explaining these patterns in color and lightness perception—and likewise for cases in hearing, touch, and so on. A proponent of the pluralist approach may, for example, have to posit representations of surface luminance in order to explain perspectival variance and similarity in color and lightness perception.⁹

The pluralist cannot give a unified explanation of the perspectival character of perception. According to the pluralist approach, perception is perspectival because of what relational properties one perceptually represents; but the relevant properties are fundamentally different for different forms of perception. Visual angle and surface luminance, for instance, do not form a natural kind. So it cannot be in virtue of representing a common kind of property that these representations form a unified psychological kind. Substantially different sensory cues and computations would be responsible for forming representations of visual angles than would be responsible for forming representations of surface luminance, and these representations would have substantially different kinds of influences on other psychological capacities.

⁹A surface's *luminance* is the total light energy reflected by that surface—a product of the surface's reflectance and the way it is illuminated.

The pluralist cannot say what underlying nature these representations share.

The disunified character of the pluralist approach is unsatisfying. It is striking that while there are many differences between spatial perception and color perception, they both exhibit patterns of variance and similarity. Perspectival representations seem to constitute a robust psychological kind. The pluralist approach purports to explain different cases of perspectival variance and similarity, but it does not offer a systematic, common account of these cases. All things being equal, we should look for a more unified, systematic account.¹⁰

4.3 The Perspectival Properties Approach

In contrast to the pluralist approach, what I call the *perspectival properties approach* offers a unified account of the perspectival character of perception. This approach posits the existence of a set of distinctive *perspectival properties* and holds that all perception involves the representation of such properties. Perspectival properties are supposed to constitute a unified natural kind of relational property and, correspondingly, representations of these properties are supposed to constitute a unified psychological kind.¹¹ Unlike the sorts of representations

¹⁰I have emphasized the disunified nature of the pluralist approach. There are other worries about attempts to explain the perspectival character of shape and color perception in terms of representations of visual angle and surface luminance, respectively. In the first place, it may be that while visual angle and surface luminance are sensorily registered, they are not represented in perception (see Burge, 2010a). A separate worry is that representations of visual angle and surface luminance, insofar as they are accurate, would not account for the patterns of perspectival variance and similarity that are found in perception. For example, Hill and Bennett (2008) argue that perspectivally similar shape representations are often not of surfaces that have similar visual angles. While these concerns are important, I have set them to the side in order to focus on broader explanatory issues.

¹¹Examples of the perspectival properties approach can be found in in (Huemer, 2001; Noë, 2004; Schellenberg, 2008; Cohen, 2010; Brogaard, 2010; Hill, 2014). Tye is sometimes grouped together with proponents of the perspectival properties approach. However, since Tye explains perspectival phenomena in visual spatial perception in terms of representations of visual angles while explaining perspectival phenomena in color perception

to which the pluralist appeals, representations of perspectival properties do not resemble the kinds of representations that ordinarily figure into empirical models of perception. In the next section, I will argue that the perspectival properties approach is not a tenable empirical hypothesis and that there is good reason that perceptual psychology standardly does not posit representations of perspectival properties.

Philosophers have characterized perspectival properties in different ways. Perspectival properties have been characterized, for example, in terms of the optical projections that objects have or else in terms of the perceptual states that objects are disposed to cause (see, respectively, Noë, 2004; Cohen, 2010). Different versions of the perspectival properties approach hold that perceptual representations of non-perspectival properties depend either epistemically, computationally, or constitutively on perceptual representations of perspectival properties (see, respectively, Schellenberg, 2008; Cohen, 2010; Noë, 2004).

In order to evaluate formulations of the perspectival properties approach empirically, one must have some conception of the conditions under which something instantiates one perspectival property or another. One needs at least a tentative account of the identity conditions of determinate perspectival properties in order to ascertain whether a representation of a perspectival property is accurate. And knowing whether a perceptual representation of a perspectival property is accurate or not is critical for developing and testing computational theories of how representations of perspectival properties are generated and employed. The claim that we represent *perspectival ellipticality*, for example, and the corresponding question of how one's visual system computes a representation of *perspectival ellipticality* are not well defined until

in terms of representations of illumination, I think it is better to classify him as a pluralist.

there is at least a preliminary conception of what *perspectival ellipticality* is.

For purposes of illustration, I will describe a *projective* model of perspectival properties. This model will help to illustrate the advantages and disadvantages of the approach. I will say that an object instantiates a particular *projective property* by virtue of projecting a certain pattern onto a projection plane relative to a viewpoint. To be *projectively elliptical* relative to a projection plane and viewpoint is to project an elliptical pattern onto the projection plane relative to that viewpoint. I will assume that the perspectival properties approach only claims that one represents items as having projective properties that correspond to the projected properties registered by the perceptual system. In the case of vision, the relevant definitions of “projection,” “projection plane,” and “viewpoint” will be specified by reference to laws of optics and the anatomy of the eye.¹² Let us think of the “projection plane” and “viewpoint” as (possibly empty) regions of physical space. So an object’s projected image can be specified as a cross-section, at a projection plane, of the light rays that the object reflects and that converge at the specified viewpoint. In canonical form:

Projective Property: An object x is projectively F relative to a viewpoint v and a projection plane p iff the optical projection of x onto p relative to v has the property of being F .

As an example, consider the circular disk in Figure 4.2a. This disk is projectively elliptical with respect to the illustrated projection plane and viewpoint. The ellipse in Figure 4.2b is

¹²For concreteness, I am characterizing projective properties in terms of optical projections at the surface of the retina. Projective properties could alternatively be specified, for example, with respect to the “Cyclopean Eye” located midway between the two actual eyes. The choice of the viewpoint and projection plane will not matter to the current discussion.

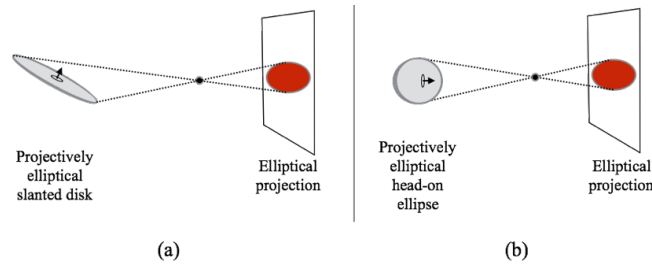


Figure 4.2: The circular disk in (a) and the ellipse in (b) are both *projectively elliptical* relative to their respective viewpoints and projection planes.

also projectively elliptical with respect to a projection plane and viewpoint.¹³

I will assume that objects have projective properties. One should be careful not to confuse the *project-ive* property with the property of the projected image (the *project-ed* property). The circular surface of a disk is *projectively elliptical* relative to the appropriate viewpoint; the pattern that the disk’s surface projects to the projection plane is simply *elliptical*. Projective shape, as I have defined it, is a property of the distal object, not of that object’s projection. One should also be careful to distinguish between the projective properties that a thing instantiates and the projective properties that a perceiver represents. While an object will simultaneously instantiate an enormous number of projective shapes, the perspectival properties approach only claims that we perceptually represent things as having the projective properties that correspond to the projected stimuli that our sensory systems are registering at a given time.

The projective characterization of perspectival properties must be treated cautiously if it is to be presented in the best light. According to the present account, the coin is perspectivally

¹³It will be convenient to suppress reference to viewpoints and projection planes and just say that an object is “projectively-*F*,” full stop, when there is some salient viewpoint v and projection plane p (as determined, for example, by the perceiver’s eye or visual system) such that the projection of x onto p relative to v has the property F . Two objects can share projective properties, in this sense, relative to different viewpoints and projection planes. For example, the disk and the ellipse in Figure 4.2 are both *projectively elliptical*, full stop.

elliptical if and only if its projection on a plane relative to a viewpoint is elliptical. It might be tempting to express the fact that a subject is representing the coin as *perspectively elliptical* by saying that the subject represents it as *having an elliptical optical projection on a plane relative to a viewpoint*. But in order to represent the coin as *having an elliptical optical projection on a plane relative to a viewpoint*, one would seemingly have to be able to perceptually represent points in space as *viewpoints*, regions in space as *projection planes*, and one would have to perceptually represent the rules of optical projection. While one may perceptually represent a point in space (for example, a point midway between the eyes) as one's viewpoint, there is little evidence to support the claim that one can perceptually represent as of projection planes or the rules of optical projection. The perspectival properties approach loses some plausibility if it is committed to one's representing these things. But it need not have this commitment.

To see how the perspectival properties approach can avoid such commitments, we should distinguish between representing a perspectival property and representing its identity conditions. The identity conditions for a perspectival property are specified by our theory. We do not need to represent a property's identity conditions in order to represent that property. So, while the perspectival properties approach holds that we represent the slanted coin as *projectively elliptical*, it need not be committed to the implausible view that we represent the coin as *having an elliptical optical projection on a plane relative to a viewpoint*.¹⁴

¹⁴The same points arise with non-projective characterizations of perspectival properties. For example, Cohen gives a dispositional, rather than projective, account of perspectival properties, and calls them "perceptual state dispositions." He writes that when we look at the slanted coin, the visual system "represents the distal item as bearing this perceptual state disposition: *disposed to generate in us an instance of the type of perceptual state we undergo when perceiving an ellipse straight on*" (Cohen, 2010, p. 110). Cohen's formulation suggests that the perceptual system represents as of *perceptual states, dispositions, causal generation*, and so on. But there is no independent empirical reason to think that the perceptual system can represent as of such properties and relations (for a related discussion, see Burge, 1991).

The unity of the perspectival properties approach can be brought out by looking at the difference in explanatory power between projective properties and visual angles. These two sorts of properties are often treated interchangeably (see, for example, Huemer, 2001; Bennett, 2009). But the concept of a projective property has a generality that the concept of visual angle does not. For any property of a projection, we can define a corresponding projective property. The concept of projective ellipticality is defined by reference to the ellipticality of an image on a projection plane. By turning to other features of a projected image, one can extend the notion of a projective property from geometrical cases, such as projective shape, to non-geometrical cases such as projective color. Projective color can be specified in terms of color, or some related property of light, at a projection plane. With a general enough concept of projection we can even define non-visual projective properties, such as the projective loudness of a sound. By contrast, the notion of a visual angle cannot be extended in any natural way to account for perspectival color or loudness perception. Because we can plausibly identify projective properties for all forms of perception, and since these properties all have the same canonical nature, one could offer a unified theory of the perspectival character of perception by appealing to representations of projective properties.

4.4 The Argument from Functional Redundancy

We now have two candidate approaches to explaining the perspectival character of perception. The pluralist approach is unsatisfying because it does not offer a unified account of the perspectival character of perception. The perspectival properties approach, by contrast, offers a

unified alternative. I will now argue that the perspectival properties approach violates a central commitment of empirical psychology, since the perceptual representations that it posits are functionally redundant. For concreteness, I will focus on a projective characterization of perspectival properties.

Moving forward, I will assume that our perceptual capacities depend on information-processing operations that generate representations of how the world is on the basis of sensory stimuli. This assumption is at the heart of what is now the standard approach to studying perception in psychology (see Marr, 1982; Palmer and Kimchi, 1986). A central commitment of this *information-processing approach* in psychology is that all psychological states make some contribution or other to information processing. For example, psychological states may carry new information or content, contribute to the production of new information or content, or help maintain information or content that is already available. From the perspective of the information-processing approach, we should not posit functionally redundant psychological states—states that play no role in information processing. This does not mean that we should never posit psychological states that play less than optimal roles in information processing. Nor does it mean that we should never posit states whose information-processing roles are not in general conducive to the survival and reproduction of the perceiver. Finally, it does not mean that we should never posit multiple psychological states that have the same or similar contents. Rather, the claim is that we should not posit psychological states that would not make any identifiable contribution whatsoever in psychological information processing. A fundamental assumption of contemporary psychology is that perceptual states must have something to do—whether or not they do it well and whether or not their doing it is good for the organism’s

well-being.

The perspectival properties approach claims that we perceptually represent perspectival properties. For this to be a tenable hypotheses, from the perspective of the information-processing paradigm in psychology, representations of perspectival properties would have to make some identifiable contribution to perceptual information processing. In fact, some proponents of the perspectival properties approach hold that representations of perspectival properties have an important role to play in perceptual processing (see, for example, Cohen, 2010). They maintain that the perceptual system computes representations of non-perspectival properties partly on the basis of prior representations of perspectival properties. On this account, when we view the slanted coin the visual system first registers an elliptical pattern on the retina, produces a representation as of a perspectively elliptical surface, and then on the basis of this transitional representation produces representation as of a circular surface. The idea is that representations of perspectival properties are stepping stones on the way to representing non-perspectival properties.

However, the claim that representations of perspectival properties are stepping stones to representing properties like size, shape, and color is untenable. I will argue that the perceptual system is set up so that it can go from registering the sensory stimulus to representing non-perspectival properties such as shape without any intermediate representations of non-perspectival properties. The regularity that objects instantiate the relevant perspectival properties is reflected in the architecture of the perceptual system, and so it does no good to also represent particular objects as instantiating those properties on particular occasions. More generally, there is no identifiable contribution that representations of perspectival properties

would make to perceptual processing.

My argument rests on a standard understanding within perceptual psychology of how the perceptual system takes advantage of environmental regularities (see Ullman, 1979; Marr, 1982; Hoffman and Richards, 1984; Shepard, 2001; Maloney, 2003; Pylyshyn, 2003). The task of perception is to represent the distal stimuli that gave rise to the proximal impacts on our sensory receptors. In principle, infinitely many distal features could have given rise to a given proximal stimulus. The task of perception is tractable, however, so long as only a subset of the distal features that *could* in principle give rise to a particular proximal stimulus *do* give rise to it and if there are regularities constraining which distal features give rise to which proximal stimuli. Evidently, the perceptual system takes advantage of this fact. The perceptual system operates as if certain kinds of regularities—called “natural constraints”—govern the way distal features give rise to proximal stimuli. Among these natural constraints are the regularities that light travels in straight lines; that light comes from above; that most points lie on rigid surfaces; for some animals, that the period of the Earth’s rotation is about twenty-four hours; and so on. Perception fails to be veridical when such regularities do not obtain.

Perceptual systems operate as if some (but perhaps not all) of these environmental regularities or natural constraints are constants that hold always and everywhere. For example, the visual system operates as if all light travels in straight lines. The standard view of information-processing theories is that the perceptual system treats these regularities as constants in virtue of the way its *architecture* is set up, where the architecture of the perceptual system consists in the information-processing operations that it carries out. Through either evolution or in-

dividual development within an environment that satisfies certain regularities, the procedures carried out by one's perceptual system produce, as a matter of course, representations that would be approximately veridical were those regularities to obtain. I will paraphrase this situation by saying that these natural constraints are "reflected" in the processing architecture of the perceptual system—in the way the information-processing operations in the perceptual system make certain perceptual outputs possible (or likely), and others impossible (or unlikely), for a given sensory input.¹⁵ To be clear, when I say that the architecture "reflects" a natural constraint I do not mean that the constraint is specified in the content of some representation in the system. Rather, the system is just set up to operate as if the constraint holds (see Kubovy and Epstein, 2001; Pylyshyn, 2003; Burge, 2010a).

The idea that certain regularities may be "reflected" in the processing architecture of a system may be clearer with a rough analogy. Suppose one wants to build a *times-seven* device that will multiply a given number by seven. Since the multiplier, seven, is always the same, the simplest way to build the device is to "hardwire" or "gear" it so that for any input representation of a number it automatically returns a representation of that number times seven. For example, suppose we represent positive numbers by the number of rotations that a gear makes. Our device may consist of two gears having the appropriate ratio of teeth so that every full rotation of the larger input gear translates into seven full rotations of the smaller output gear. The multiplier, seven, is reflected in the relative sizes of the gears. It is important to note that the device's ability to multiply numbers by seven does not depend on

¹⁵It is worth noting that whether a natural constraint is reflected in the architecture of the perceptual system is independent of whether the relevant aspects of the system's architecture are innate or learned, fixed or changeable.

its occasioning a representation of the multiplier in the turns of a particular gear.

Here is the first central point of my argument against the claim that we represent perspectival properties. The architecture of the visual system reflects the natural constraint that objects instantiate projective properties corresponding to the patterns registered on the retina. In fact, some of the most basic natural constraints on visual perception come from the laws of optics, or some approximation of them. It is a widespread commitment of empirical models in perceptual psychology that the architecture of our visual system is set up to operate as if something like the laws of optics always govern the formation of the inputs to vision.¹⁶ So, given an elliptical retinal image, for example, a normal visual system could only output representations of shapes (circles and ellipses) that could have projected that input at some orientation or other. A normally functioning visual system could not return a representation as of a square in response to an elliptical pattern on the retina.

Now for the second central point of the argument. There is no functional benefit to representing what is already reflected in a system's architecture. The visual system does not, for example, need to represent the projective ellipticality of the slanted coin in order to represent that coin's circularity, since the visual system is already constrained by its architecture to only generate representations of those shapes that could have projected the elliptical pattern at the eye. The system can go from registering the proximal stimulus to representing the

¹⁶Indeed, perceptual psychologists often are interested in more substantial natural constraints that either entail or rely on the laws of optics—for example, the constraint that collinear lines in the retinal image are only projected by collinear contours in the scene, or that points of deep concavity in the retinal image are only projected by points at which one object either intersects or occludes another. Psychologists typically suppose that the architecture of the visual system reflects even these natural constraints, which are in fact less universal than the laws of optics (see, for example, Hoffman and Richards, 1984).

non-projective properties of the distal stimulus, without any intermediate representations of projective properties.

Consider, by analogy, our times-seven device. Since the multiplier, seven, is reflected in the relative sizes of the input and output gears, there is no functional benefit to including a separate gear whose rotations represent the multiplier. The representation of the multiplier is not needed as an intermediate step in generating the right output. Likewise, if the processing architecture of the perceptual system reflects the constraint that perceived objects project the registered proximal stimuli, then there is no functional benefit to also tokening states that represent particular objects as projecting the registered stimuli. Tokening such states would not contribute anything over and above what is already guaranteed by the architecture of the perceptual system.

Since representations of projective properties would be functionally redundant, we should not posit them. The dominant information-processing paradigm in psychology assumes that psychological states must make some sort of contribution to information processing. But since the architecture of the perceptual system already reflects the constraint that the objects of perception project the registered sensory inputs, perceptual representations of objects as instantiating projective properties would not contribute anything to perceptual information processing. The perspectival properties approach violates a foundational commitment of perceptual psychology by positing representations that would make no identifiable contribution to perceptual information processing.

Non-projective variants of the perspectival properties approach face versions of this same

argument. Cohen, for example, offers a dispositional account on which to be *perspectively elliptical* is to be *disposed to generate in us an instance of the type of perceptual state we undergo when perceiving an ellipse straight on*. But the architecture of the perceptual system already reflects the constraint that the objects of perception are disposed to be the distal causes of the system's states. It is natural that the perceptual should reflect this constraint because, plausibly, the primary and unchanging function of the visual system is to represent the distal causes of its states. If the perceptual system reflects this constraint, then there is no reason for the system to occasion representations of particular objects as *disposed to cause such-and-such states of the perceptual system* (see also Burge, 1991).

The perspectival properties approach promises a unified account of the perspectival character of perception, but at the expense of positing perceptual representations that would serve no role in perceptual information processing. Achieving a unified account is not a sufficient motivation for positing such states.

4.5 The Structures of Perspectival Representations

The view that the perspectival character of perception depends on our perceptually representing certain relational properties standardly takes the form of either a disunified, pluralist account or an account that posits functionally redundant psychological states. I will now recommend a different sort of approach that has neither of these drawbacks. According to what I call the *structural approach*, the central difference between perspectively variant representations is that they are structured in different ways from their parts, while the central commonality

between perspectively similar representations is that they are structured in similar ways from their parts. After introducing the notion that perceptual states have part-whole structure, I will illustrate how differences and similarities in this kind of structure can account for perspectival variance and similarity. I will then argue that what unifies the different systems of perspectival representation is the way in which, in those systems, representations of properties such as size, shape, and color are structurally interlocked with representations of properties such as distance, orientation, and illumination.

A basic assumption in perceptual psychology is that representational perceptual states—the vehicles of perceptual content—have part-whole structure. As I described in Chapter 1, perceptual states can, so to speak, be “made of” or arranged from other constituent perceptual states, much as sentences are made of component words and maps are made of colored regions. For example, my representation of the tabletop as brown and rectangular has as constituent parts my representation as of an instance of brownness and my representation as of an instance of rectangularity. The structure of a representation consists in the way its constituent parts are combined. The information-processing architecture of a system is intimately related to the structure of the representations over which the system operates. Some of the core processes in perception involve building up, breaking down, traversing through, or comparing the structures of representations.¹⁷

The structure of a perceptual state is an important representational feature of that state.

¹⁷For classic discussions of how psychological representations are structured and how their structure relates to information-processing architectures, see (Palmer, 1978; Kosslyn, 1980; Marr, 1982; Pylyshyn, 1984).

Psychologists sometimes use the term “format” to refer generally to the available ways of structuring representations in a system. Sentences and maps have different formats, in this sense, while a map of South America and a map of Africa may have the same format but different structures.

In general, the structure of a representation plays an essential role in its representing what it does. I argued in Chapter 2 that what something represents normally depends on what its constituents represent and how they are combined. What state of affairs a sentence represents, for example, depends on the individuals and properties that the words in that sentence denote and how those words are put together. What a map depicts depends on what the colors in that map depict and how the colored regions of the map are arranged. On the other hand, a representation's structure is not fully determined by what it represents. A map and a (very long) sentence could represent all the same locations and relations among all the same landmarks. Yet these representations have different structures, being made up in different ways from different sets of primitive representations.

How might the structure of a representation account for its perspectival character? Consider paintings made in a realistic style. Such paintings are paradigmatic cases of perspectival representations. Intuitively, such paintings exhibit perspectival variance and similarity because of how they are structured—how they are composed from colored marks on a two-dimensional surface. The pattern you paint in order to depict a slanted coin has to be different from the pattern you use to depict a head-on coin (perspectival variance) and will be similar to the pattern you paint to depict a head-on ellipse (perspectival similarity). The paint you use to depict a shaded white surface has to be different from the paint you use to depict an unshaded white surface (perspectival variance) and similar to the paint you use to depict an unshaded gray surface (perspectival similarity). There is no pressing need to hypothesize that paintings depict visual angles, surface luminance, or perspectival properties. The perspectival character of paintings is a product of their structure.

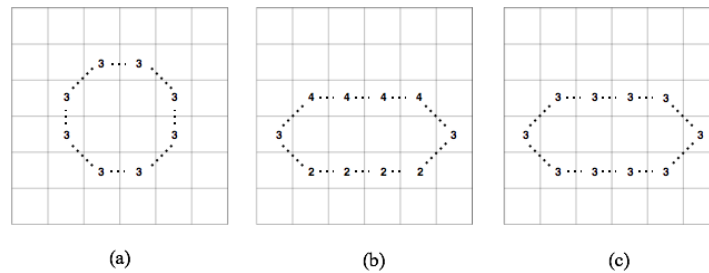


Figure 4.3: Rough illustrations of array representations as of (a) a head-on circular surface, (b) a slanted circular surface, and (c) a head-on elliptical surface. The primitive constituents are numerals that represent contour segments as being located at particular distances from the viewpoint. The direction in which one represents a contour segment as located is a function of where the numeral is placed in the array. Dotted lines indicate which sets of primitives integrate into complex representations. These complex representations attribute features like size and shape to whole surfaces, depending on what distances their primitive parts represent and how those parts are placed in the array.

We can turn to some empirical hypotheses about the structure of perceptual representations to illustrate how the structures of perceptual states accounts for the perspectival character of perception—though the structural approach need not be tied to these specific hypotheses. I argued in Chapter 3 that a core class of perceptual representations are structured like arrays, analogous to those in Figure 4.3 (see also Marr, 1982; Evans, 1985a; Peacocke, 1992; Burge, 2014).¹⁸ Remember that to say a perceptual representation is array-like is not to say that it is literally laid out in space, nor does it suggest that we have some “inner eye” that views and interprets the patterns in an “inner array” (see Block, 1983). Rather, the view is that we can best explain how certain perceptual representations are processed and how they can combine with each other in terms of what cells in an array the constituents of those representations occupy. The constituents of array-like perceptual representations are perceptual states that represent

¹⁸The literature on *mental imagery* contains many discussions about the concept of an array-like representation (see Kosslyn, 1980; Pinker, 1988; Tye, 1991).

patches and edges of surfaces as having certain orientations and distances from the perceiver. The cell or address that a constituent representation has in the representational array, like a pixel's place in a digital photograph, corresponds to the line of sight along which that constituent represents a patch of surface—with adjacent cells corresponding to adjacent lines of sight. An array-like representation of a surface is a combination of representations in these cells. It is plausible that array-like perceptual representations underlie important aspects of conscious visual experience (for arguments to this effect, see Jackendoff, 1987; Prinz, 2007; Bennett, 2012).

If perceptual representations of surfaces are array-like, then the representation of the head-on coin's circularity must have a different structure than the representation of the slanted coin's circularity. The representations of the head-on coin and the slanted coin must be organized over different cells in the representational array, since the head-on coin occupies different lines of sight than the slanted coin (compare with Figure 4.3a and b). This structural difference accounts for perspectival variance in one's perception of the coin.

On the other hand, the representation of the slanted coin is structurally similar to the representation of the head-on ellipse. To be sure, the two representations will have different constituents. The representation of the slanted coin will be organized from representations as of slanted surface patches, whereas the representation of the head-on ellipse will be organized from representations as of head-on surface patches. It is because the perceptual states that represent the slanted coin and the head-on ellipse have different sets of constituents that the one state represents the coin as slanted and circular and the other state represents the ellipse as head-on and elliptical. However, while the two perceptual states represent different shapes and

are arranged from different set of constituents, those sets of constituents are organized over the same pattern of cells in the representational array, since we see the points on the slanted coin and the head-on ellipse along all the same lines of sight (compare with Figure 4.3b and c). This structural similarity accounts for the perspectival similarity of our perceptions of the slanted coin and the head-on ellipse.¹⁹

While the structures of the representations involved in lightness and color perception are in some ways more difficult to describe, there is good reason to think that they are responsible for the perspectival phenomena we find there. Plausibly, how one represents a surface's lightness and color depends on how one represents that surface as illuminated.²⁰ Light at the retina is the product of both the material properties of the surfaces in front of us and of the way those surfaces are illuminated. A task of color vision is to decompose the light at the retina into material lightness and color components and an illumination component. The visual system seems to reflect the regularity that the represented material components and the represented illumination component jointly account for the light registered at the eye. This

¹⁹Tye (1996) draws a similar connection between the array-like structure of our perceptual representations and the impression that a farther tree looks different and, in some sense "smaller," than a nearer tree of the same size. Tye suggests that array-like perceptual states represent visual angles. On Tye's final analysis, our perception of the farther tree is perspectivally different from our perception of the nearer, same-sized tree because we attribute different visual angles to the trees; the distant tree is perspectivally similar to a nearer, smaller tree because we attribute the same visual angles to these trees. In contrast to Tye, I do not claim that array-like perceptual states represent visual angles. I believe that even if a representation's place in the representational array corresponds to the line of sight along which a point in the scene is represented, combinations of these representations need not represent as of the visual angles between those lines of sight. More importantly, however, I claim that a unified account of perspectival variance and similarity should center on the *structure* of our perceptual states and not on what properties, such as visual angle, those states might represent.

²⁰While I think it is plausible that we represent how surfaces are illuminated there is no current consensus on the matter. Many models of lightness and color perception assume that the visual system merely filters out the effect of illumination without representing that effect. If we do not represent something like how surfaces are illuminated, then the structural approach will have to be pursued along different lines than I suggest here (for some discussion of these issues, see Maloney and Yang, 2003; Gilchrist, 2006; Kingdom, 2011).

means that represented lightness and color and represented illumination have a complementary relationship. If, given the light at the eye, you can represent a surface as a well-lit gray or else as a shaded white, you could not normally represent that surface on that occasion as a shaded gray or a well-lit white. This regularity may well be embodied in the very structure of representations of lightness and color. One can then think of the registered light at the eye as setting a structural constraint on how representations of lightness and color can combine with representations of illumination. For example, a representation of a surface as shaded and white or a representation of the surface as unshaded and gray may both be possible under a particular structural constraint, while a representation of the surface as shaded and gray would not be possible under that constraint. If this is right, then one's representations of the shaded white surface and of the unshaded white surface will differ because they are subject to different structural constraints. And one's representations of the shaded white surface and of the unshaded gray surface will be similar because they fall under a similar structural constraint (see also Mausfeld, 2003).

I have been proposing that the perspectival character of size, shape, lightness, and color perception depends on the structures of the representations involved. I now want to suggest that all types of perspectival variance and similarity result from a common kind of structural characteristic: in all perspectival forms of perception, representations as of non-relational properties (for example, size, shape, surface lightness, color) and representations of relational properties that characterize how things relate to us and to their surroundings (for example, distance, direction, orientation, illumination) are *structurally interdependent*. Following the discussion in Chapter 1, I will say that one type of representation α structurally depends on

another type of representation β if sub-types of β constrain the sub-types of α with which they can combine. For example, shape representations structurally depend on orientation representations if representations of particular slants constrain the representations of circularity (or rectangularity, or what have you) with which they can combine.

Paintings offer good examples of the type of structural interdependency that I have in mind. The way paintings are put together by combining colored marks on a two-dimensional surface requires that the types of marks you use to depict shape and lightness/color intrinsically depend on the types of marks you use to depict orientation and illumination, respectively. In principle, there are only certain ways you can depict the shape of a surface given that you are depicting the surface at a certain orientation, and *vice versa*. This structural interdependency gives rise to perspectival variance and similarity in paintings. Because the depiction of orientation structurally constrains the depiction of shape, you cannot paint a slanted disk the same way that you paint a head-on disk and you must paint a slanted disk in a similar way as you paint a head-on ellipse. Likewise, how you depict a surface's color depends on how you depict the way it is illuminated, and *vice versa*. As a result of this interdependency, you cannot paint a shaded white surface the way you paint an unshaded white surface and the way you paint a shaded white surface must be similar to the way you paint an unshaded gray surface.

The sort of structural interdependencies in painting show up throughout perception. For example, what structure an array-like representation of a circular surface can have depends on how one represents the surface as oriented. Conversely, how we represent a surface as oriented depends in part on the structure of our array-like representation of that surface. In the case of color perception, it is plausible that one's representation of the way a surface is

illuminated constrains what color representations are structurally possible, and *vice versa*. Perspectival variance and similarity arise out of the specific ways in which representations of non-relational properties such as size, shape, lightness, and color and representations of relational properties such as distance, orientation, and illumination, are interdependent.

In my objection to the perspectival properties view, I argued that there is no need to represent the objects of perception as producing the registered proximal stimuli because that they do so is reflected in the architecture of the perceptual system. This point is not incidental to the current proposal. If certain constraints are reflected in the very architecture of the system, then we have reason to expect those constraints to be embodied in the structures of the representations in which the system traffics. For example, if the architecture of the visual system reflects the constraint that two properties, such as *blue* and *yellow*, cannot co-occur at the same point, it is reasonable to infer that the *representation* as of a point as *blue* and the *representation* as of that point as *yellow* are structurally incompatible—the perceptual states cannot be combined to form a representation of a point as *blue and yellow*.

The structural interdependencies between perceptual representations of size and distance, shape and orientation, lightness/color and illumination, are rooted in the way the perceptual system must disentangle these pairs of properties. The perceptual system capitalizes on natural constraints in order to distinguish between distal non-relational properties such as shape and color and distal relational properties such as surface orientation and illumination, which have been confounded in the proximal stimulus. As I emphasized in the previous section, the architecture of the perceptual system reflects constraints on how those features must go together to produce the proximal stimulus. While I argued that we do not represent those constraints,

they nevertheless seem to shape the structures of our perceptual representations, so that there are only certain ways of structuring representations of non-relational properties given how one represents the relevant relational properties. For example, the array-like structure of our perceptual representations of surfaces embodies the constraint that, given an elliptical pattern on the retina, you must be looking at either a slanted circle or a head-on ellipse (see Burge, 2014, p. 494–5). An elliptical pattern on the retina must have been produced by a surface that occupies certain lines of sight. So, under normal conditions the representations of points on that surface will fill an elliptical pattern of cells in the representational array. Depending on how one represents the orientations of (or distances to) those points, the combination of the representations that fill those cells must represent either a slanted circle or a head-on ellipse.

The guiding idea behind the structural approach is that the perspectival character of representations depends on the structural features of those representations—on the nature of their representational ingredients and the ways those ingredients are combined. Perspectival variance and similarity result from differences and similarities in the way perceptual representations are structured from their parts. There is a common thread behind what makes different forms of perception perspectival. In each case, representations of non-relational properties such as size, shape, lightness, and color and representations of relational properties such as distance, orientation, and illumination are structurally interdependent. It is no accident that these structural interdependencies show up throughout perception. These interdependencies correspond to the way the perceptual system’s architecture reflects natural constraints on how proximal stimuli confound distal properties. While the structural approach calls for more development, it promises to offer an empirically plausible, unified explanation of the perspectival

character of perception.

4.6 Conclusion

Many have held that perception is perspectival because we perceive things as having certain relational properties that correspond to the varying ways those things look relative to different conditions of observation. The pluralist version of this view explains the perspectival character of spatial perception in terms of the perception of one kind of property, such as visual angle, while explaining the perspectival character of color perception in terms of another kind of property, such as surface luminance. This approach is unsatisfying because it fails to give a systematic, unified account of the perspectival character of perception. By contrast, the perspectival properties approach offers a unified account on which perception is perspectival because we always perceive things as having perspectival properties. This approach is untenable because it posits perceptual representations that would have no identifiable function in perceptual information processing.

I have proposed a different explanation of the perspectival character of perception. Perceptual representations are perspectival because of the ways they are structured from their parts. What perspectival systems of representation have in common is that their representations of non-relational properties (such as size, shape, lightness, and color) and their representations of relational properties (such as distance, orientation, illumination) are structurally interdependent. These structural interdependencies are rooted in the way perception works to disentangle the contributions that those properties make in producing sensory stimuli. Per-

spectival representations form an explanatorily unified psychological kind not by virtue of the properties that they represent but by virtue of how they are structured.

Bibliography

- Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, 262(5142):2042–2044.
- Adelson, E. H. (2000). Lightness perception and lightness illusions. In Gazzaniga, M., editor, *The New Cognitive Neurosciences*, pages 339–351. MIT Press, Cambridge, MA, 2nd edition.
- Anderson, B. L. and Khang, B.-G. (2010). The role of scission in the perception of color and opacity. *Journal of Vision*, 10(5):26, 1–16.
- Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, 85(4):249–277.
- Ankrum, C. and Palmer, J. (1991). Memory for objects and parts. *Perception & Psychophysics*, 50(2):141–156.
- Arguin, M. and Saumier, D. (2004). Independent processing of parts and of their spatial organization in complex visual objects. *Psychological Science*, 15(9):629–633.
- Ashby, F. G., Prinzmetal, W., Ivry, R., and Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, 103(1):165–192.

- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193.
- Attneave, F. (1968). Triangles as ambiguous figures. *The American Journal of Psychology*, 81(3):447–453.
- Awh, E. and Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2):834–846.
- Aydede, M. (1997). Language of thought: The connectionist contribution. *Minds and Machines*, 7(1):57–101.
- Balcerak Jackson, B. (2017). Structural entailment and semantic natural kinds. *Linguistics and Philosophy*, 40(3):207–237.
- Balog, K. (2009). Jerry Fodor on non-conceptual content. *Synthese*, 170(2):311–320.
- Barenholtz, E. and Feldman, J. (2003). Visual comparisons within and between object parts: evidence for a single-part superiority effect. *Vision Research*, 43(15):1655–1666.
- Baylis, G. C. and Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3):451–470.
- Beall, J. and Restall, G. (2006). *Logical Pluralism*. Clarendon Press, Oxford.
- Behrmann, M., Zemel, R. S., and Mozer, M. C. (1998). Object-based attention and occlusion:

- Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4):1011–1036.
- Beller, H. (1971). Priming: Effects of advance information on matching. *Journal of Experimental Psychology*, 87(2):176–182.
- Bennett, D. J. (2009). Varieties of visual perspectives. *Philosophical Psychology*, 22(3):329–352.
- Bennett, D. J. (2012). Seeing shape: Shape appearances and shape constancy. *The British Journal for the Philosophy of Science*, 63(3):487–518.
- Berwick, R. C. and Weinberg, A. S. (1983). The role of grammars in models of language use. *Cognition*, 13:1–61.
- Bhatt, R. S. and Bertin, E. (2001). Pictorial cues and three-dimensional information processing in early infancy. *Journal of Experimental Child Psychology*, 80:315–332.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147.
- Biederman, I. and Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology*, 23(3):393–419.
- Blaser, E., Pylyshyn, Z. W., and Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, 408:196–199.

- Block, N. (1980). Troubles with functionalism. In Block, N., editor, *Readings in Philosophy of Psychology*, volume 1, pages 268–305. Harvard University Press, Cambridge, MA.
- Block, N. (1983). Mental pictures and cognitive science. *The Philosophical Review*, 92(4):499–541.
- Bonnay, D. (2008). Logicality and invariance. *The Bulletin of Symbolic Logic*, 14(1):29–68.
- Brogaard, B. (2010). Strong representationalism and centered content. *Philosophical Studies*, 151(3):373–392.
- Brunswik, E. and Kamiya, J. (1953). Ecological cue-validity of ‘proximity’ and of other gestalt factors. *The American Journal of Psychology*, 66(1):20–32.
- Burge, T. (1986). Individualism and psychology. *The Philosophical Review*, 95(1):3–45.
- Burge, T. (1991). Vision and intentional content. In Lepore, E. and van Gulick, R., editors, *John Searle and His Critics*, chapter 14, pages 195–213. Blackwell, Cambridge, MA.
- Burge, T. (2005). Postscript to “Frege and the hierarchy”. In *Truth, Thought, Reason: Essays on Frege*, pages 167–210. Clarendon Press, Oxford.
- Burge, T. (2007). Intentional properties and causation. In *Foundations of Mind: Philosophical Essays, Volume 2*, chapter 15, pages 335–343. Clarendon Press, Oxford.
- Burge, T. (2009). Five theses on de re states and attitudes. In Almog, J. and Leonardi, P., editors, *The Philosophy of David Kaplan*, pages 233–316. Oxford University Press.
- Burge, T. (2010a). *Origins of Objectivity*. Oxford University Press, New York.

- Burge, T. (2010b). Origins of perception. *Disputatio*, IV(29):1–38.
- Burge, T. (2014). Reply to Rescorla and Peacocke: Perceptual content in light of perceptual constancies and biological constraints. *Philosophy and Phenomenological Research*, 88(2):485–501.
- Byrne, A. and Hilbert, D. R. (2003). Color realism and color science. *Behavioral and Brain Sciences*, 26(1):3–64.
- Camp, E. (2004). The generality constraint and categorial restrictions. *The Philosophical Quarterly*, 54(215):209–231.
- Camp, E. (2007). Thinking with maps. *Philosophical Perspectives*, 21(1):145–182.
- Camp, E. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research*, 78(2):275–311.
- Cant, J. S., Large, M.-E., McCall, L., and Goodale, M. A. (2008). Independent processing of form, colour, and texture in object perception. *Perception*, 37(1):57–78.
- Carrasco, M. and Seamon, J. G. (1996). Priming impossible figures in the object decision test: The critical importance of perceived stimulus complexity. *Psychonomic Bulletin & Review*, 3(3):344–351.
- Carruthers, P. (2004). On being simple minded. *American Philosophical Quarterly*, 41(3):205–220.

- Casati, R. and Varzi, A. C. (1999). *Parts and Places: The Structures of Spatial Representation*. The MIT Press.
- Cavina-Pratesi, C., Kentridge, R., Heywood, C., and Milner, A. (2010). Separate channels for processing form, texture, and color: Evidence from fmri adaptation and visual object agnosia. *Cerebral Cortex*, 20(10).
- Chalmers, D. J. (2004). The representational character of experience. In Leiter, B., editor, *The Future for Philosophy*, pages 153–181. Oxford University Press, New York.
- Chalmers, D. J. (2006). Perception and the fall from eden. In Szabó Gendler, T. and Hawthorne, J., editors, *Perceptual Experience*, pages 49–125. Oxford University Press, New York.
- Chase, W. G. and Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1):55–81.
- Chomsky, N. (1957). *Syntactic Structures*. Walter de Gruyter.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Clapp, L. and Duhau, L. (2011). Varieties of the generality constraint. *Manuscripta*, 34(2):397–434.
- Clark, A. (2004). Feature-placing and proto-objects. *Philosophical Psychology*, 17(4):443–469.
- Clowes, M. (1971). On seeing things. *Artificial Intelligence*, 2(1):79 – 116.

- Cohen, A. and Ivry, R. B. (1991). Density effects in conjunction search: Evidence for a coarse location mechanism of feature integration. *Journal of Experimental Psychology: Human Perception and Performance*, 17(4):891–901.
- Cohen, J. (2004). Color properties and color ascriptions: A relationalist manifesto. *The Philosophical Review*, 113(4):451–506.
- Cohen, J. (2010). Perception and computation. *Philosophical Issues*, 20:96–124.
- Cooper, L. A. and Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In Chase, W. G., editor, *Visual Information Processing*, chapter 3. Academic Press.
- Crane, T. (1990). The language of thought: No syntax without semantics. *Mind & Language*, 5(3):187–212.
- Cresswell, M. (1985). *Structured Meanings*. The MIT Press, Cambridge, MA.
- Cummins, R. (1996). Systematicity. *The Journal of Philosophy*, 93(12):591–614.
- Cummins, R. (2010). Representational specialization: The synthetic a priori revisited. In *The World in the Head*, pages 194–209. Oxford University Press, New York.
- De Winter, J. and Wagemans, J. (2006). Segmentation of object outlines into parts: a large-scale integrative study. *Cognition*, 99(3):275–325.
- Denha, S. L. and Winkler, I. (2015). Auditory perceptual organization. In Wagemans, J., editor, *The Oxford Handbook of Perceptual Organization*, pages 601–620. Oxford University Press, New York.

- Dever, J. (1999). Compositionality as methodology. *Linguistics and Philosophy*, 22:311–326.
- Devitt, M. (2006). *Ignorance of Language*. Oxford University Press, New York.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113(4):501–517.
- Egan, F. (2014). How to think about mental content. *Philosophical Studies*, 170(1):115–135.
- Elder, J. and Zucker, S. (1994). A measure of closure. *Vision Research*, 34(24):3361–3369.
- Elder, J. H. (2015). Bridging the dimensional gap: Perceptual organization of contour into two-dimensional shape. In Wagemans, J., editor, *The Oxford Handbook of Perceptual Organization*, chapter 11, pages 207–235. Oxford University Press.
- Elder, J. H. and Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2(4):324–353.
- Enns, J. T. and Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, 98(3):335.
- Etchemendy, J. (1990). *The Concept of Logical Consequence*. Harvard University Press, Cambridge, MA.
- Evans, G. (1982). *The Varieties of Reference*. Clarendon Press.
- Evans, G. (1985a). Molyneux’s question. In *Collected Papers*, pages 364–399. Oxford University Press, New York.

- Evans, G. (1985b). Semantic structure and logical form. In *Collected Papers*, pages 49–75. Clarendon Press, New York.
- Feldman, J. (1997). Regularity-based perceptual grouping. *Computational Intelligence*, 13(4):582–623.
- Feldman, J. (1999a). Does vision work? towards a semantics of perception. In Lepore, E. and Pylyshyn, Z., editors, *What is Cognitive Science?*, pages 208–229. Basil Blackwell.
- Feldman, J. (1999b). The role of objects in perceptual grouping. *Acta Psychologica*, 102(2–3):137–163.
- Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences*, 7(6):252 – 256.
- Field, H. H. (1978). Mental representation. *Erkenntnis*, 13(1):9–61.
- Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28(2):97–115.
- Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press, Cambridge, MA.
- Fodor, J. A. (1983). Methodological solipsism considered as a research strategy in cognitive psychology. In *Representations: Philosophical Essays on the Foundations of Cognitive Science*, pages 225–253. MIT Press.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Mit Press.

- Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1–2):3–71.
- Foster, D. H. (2011). Color constancy. *Vision Research*, 51(7):674–700.
- Fougnie, D. and Alvarez, G. A. (2011). Object features fail independently in visual working memory: evidence for a probabilistic feature-store model. *Journal of Vision*, 11(3):1–12.
- Frege, G. (1914). Letter to Jourdain. In Beany, M., editor, *The Frege Reader*, pages 319–321. Blackwell Publishing, Oxford.
- Frege, G. (1963). Compound thoughts. *Mind*, 72(285):1–17.
- Frisby, J. P. and Stone, J. V. (2010). *Seeing: The Computational Approach to Biological Vision*. The MIT Press, Cambridge, MA, 2nd ed. edition.
- Fu, K. S. (1974). *Syntactic Methods in Pattern Recognition*. Academic Press.
- Gallant, J. L. and Garner, W. (1988). Some effects of distance and structure on conjunction errors. *Bulletin of the Psychonomic Society*, 26(4):323–326.
- Garner, W. R. (1974). *The Processing of Information and Structure*. Lawrence Elbaum Associates, New York.
- Garrigan, P. and Kellman, P. (2011). The role of constant curvature in 2-d contour shape representations. *Perception*, 40(11):1290–1308.
- Geach, P. (1957). *Mental Acts: Their Content and Their Objects*. Routledge & Kegan Paul, London.

- Geisler, W. S., Perry, J., Super, B., and Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41(6):711–724.
- Geisler, W. S. and Super, B. J. (2000). Perceptual organization of two-dimensional patterns. *Psychological Review*, 107(4):677–708.
- Gilchrist, A. (2006). *Seeing Black and White*. Oxford University Press, New York.
- Gilchrist, A. (2015). Perceptual organization in lightness. In *Oxford Handbook of Perceptual Organization*. Oxford University Press.
- Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., and Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, 106(4):795–834.
- Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1):86–112.
- Green, E. and Quilty-Dunn, J. (2017). What is an object file? *British Journal for the Philosophy of Science*, axx055.
- Greenberg, G. J. (2011). *The Semiotic Spectrum*. PhD thesis, Rutgers University.
- Grill-Spector, K. and Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, 27:649–677.
- Guzmán, A. (1968). Decomposition of a visual scene into three-dimensional bodies. In *Pro-*

- ceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I, AFIPS '68* (Fall, part I), pages 291–304, New York, NY, USA. ACM.
- Hazeltine, R. E., Prinzmetal, W., and Elliot, K. (1997). If it's not there, where is it? Locating illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1):263–277.
- Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Wiley.
- Hering, E. (1964). *Outlines of a Theory of the Light Sense*. Harvard University Press, Cambridge, MA.
- Higginbotham, J. (1986). Linguistic theory and Davidson's program in semantics. In LePore, E., editor, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, pages 29–48. Basil Blackwell.
- Hill, C. S. (2014). The content of visual experience. In *Meaning, Mind, and Knowledge*, pages 218–236. Oxford University Press, New York.
- Hill, C. S. and Bennett, D. J. (2008). The perception of size and shape. *Philosophical Issues*, 18:294–315.
- Hinton, G. E. and Parsons, L. M. (1988). Scene-based and viewer-centered representations for comparing shapes. *Cognition*, 30(1):1–35.
- Hodges, W. (1998). Compositionality is not the problem. *Logic and Logical Philosophy*, 6:7–33.

- Hoffman, D. D. and Richards, W. A. (1984). Parts of recognition. *Cognition*, 18(1):65–96.
- Holcombe, A. O. and Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, 4(2):127–128.
- Hopp, W. (2013). No such look: Problems with the dual content theory. *Phenomenology and the Cognitive Sciences*, 12(4):813–833.
- Horn, B. K. (1974). Determining lightness from an image. *Computer Graphics and Image Processing*, 3(4):277–299.
- Huemer, M. (2001). *Skepticism and the Veil of Perception*. Rowman & Littlefield, New York.
- Huffman, D. A. (1971). Impossible objects as nonsense sentences. *Machine intelligence*, 6(1):295–323.
- Humphreys, G. W. (1983). Reference frames and shape perception. *Cognitive Psychology*, 15(2):151–196.
- Humphreys, G. W. and Quinlan, P. T. (1988). Priming effects between two-dimensional shapes. *Journal of Experimental Psychology: Human Perception and Performance*, 14(2):203–220.
- Hurvich, L. M. and Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, 64(6):384–404.
- Jackendoff, R. (1987). *Consciousness and the Computational Mind*. The MIT Press, Cambridge, MA.

- Jackendoff, R. S. (1983). *Semantics and Cognition*. The MIT Press, Cambridge, MA.
- Jagnow, R. (2012). Representationalism and the perspectival character of perceptual experience. *Philosophical Studies*, 157(2):227–249.
- Janssen, T. M. (2011). Compositionality. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, chapter 10, pages 495–553. Elsevier, 2nd edition.
- Johnson, K. (2004). On the systematicity of language and thought. *The Journal of Philosophy*, 101(3):2004.
- Johnson, K. (2015). Maps, languages, and manguages: Rival cognitive architectures? *Philosophical Psychology*, 28(6):815–836.
- Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2):175–219.
- Kaneko, H. and Uchikawa, K. (1997). Perceived angular and linear size: The role of binocular disparity and visual surround. *Perception*, 26(1):17–27.
- Kappers, A. M. and Bergmann Tiest, W. M. (2015). Tactile and haptic perceptual organization. In Wagemans, J., editor, *The Oxford Handbook of Perceptual Organization*, pages 621–638. Oxford University Press, New York.
- Kaufman, L., Kaufman, J. H., Noble, R., Edlund, S., Bai, S., and King, T. (2006). Perceptual distance and the constancy of size and stereoscopic depth. *Spatial Vision*, 19(5):439–457.

- Kellman, P. J., Garrigan, P., and Erlikhman, G. (2013). Challenges in understanding visual shape perception and representation: Bridging subsymbolic and symbolic coding. In Dickinson, S. and Pizlo, Z., editors, *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective*, pages 249–274. Springer-Verlag, London.
- Kellman, P. J. and Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23(2):141–221.
- Kimchi, R. (2000). The perceptual organization of visual objects: A microgenetic analysis. *Vision Research*, 40(10–12):1333–1347.
- Kimchi, R. (2009). Perceptual organization and visual attention. In Srinivasan, N., editor, *Progress in Brain Research*, volume 176, chapter 2, pages 15–33. Elsevier.
- Kimchi, R. (2015). The perception of hierarchical structure. In *Oxford Handbook of Perceptual Organization*, chapter 7, pages 129–149. Oxford University Press.
- Kimchi, R., Hadad, B., Behrmann, M., and Palmer, S. E. (2005). Microgenesis and ontogenesis of perceptual organization: Evidence from global and local processing of hierarchical patterns. *Psychological Science*, 16(4):282–290.
- Kimchi, R. and Hadad, B.-S. (2002). Influence of past experience on perceptual grouping. *Psychological Science*, 13(1):41–47.
- Kingdom, F. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, 51(7):652–673.

- Kitcher, P. (1988). Marr's computational theory of vision. *Philosophy of Science*, 55(1):1–24.
- Koffka, K. (1936). *Principles of Gestalt Psychology*. London: Kegan Paul, Trench, Trubner & Co., Ltd.
- Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. New York: Liveright Publishing Corporation.
- Koopman, H. (2010). Prepositions, postpositions, circumpositions, and particles. In Cinque, G. and Rizzi, L., editors, *Mapping Spatial PPs: The Cartography of Syntactic Structures*, volume 6, pages 26–73. Oxford University Press, New York.
- Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics*, 14(1):90–94.
- Kosslyn, S. M. (1980). *Image and Mind*. Harvard University Press, Cambridge, MA.
- Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review*, 88(1):46–66.
- Kosslyn, S. M., Ball, T. M., and Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 4(1):47–60.
- Kovács, I. (2000). Human development of perceptual organization. *Vision Research*, 40(10–12):1301–1310.

- Kovács, I. and 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.
- Kracht, M. (2008). The fine structure of spatial expressions. In Asbury, A., Dotlačil, J., Gehrke, B., and Nouwen, R., editors, *Syntax and Semantics of Spatial P*, pages 35–62. John Benjamins Publishing Company, Philadelphia.
- Kubovy, M. and Epstein, W. (2001). Internalization: A metaphor we can live without. *Behavioral and Brain Sciences*, 28(4):618–625.
- Land, E. H. (1977). The retinex theory of color vision. *Scientific American*, 237(6):108–128.
- Land, E. H. (1983). Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image. *Proceedings of the National Academy of Sciences*, 80:5163–5169.
- Land, E. H. and McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11.
- Lande, K. J. Objects in space. Unpublished manuscript.
- Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35(3):389–412.
- Leek, E., Reppa, I., and Arguin, M. (2005). The structure of three-dimensional object repre-

- sentations in human vision: Evidence from whole-part matching. *Journal of Experimental Psychology: Human Perception and Performance*, 31(4):668–684.
- Leek, E. C., Reppa, I., Rodriguez, E., and Arguin, M. (2009). Surface but not volumetric part structure mediates three-dimensional shape representation: Evidence from part-whole priming. *The Quarterly Journal of Experimental Psychology*, 62(4):814–830.
- Leeuwenberg, E. (1971). A perceptual coding language for visual and auditory patterns. *The American Journal of Psychology*, 84(3):307–349.
- Lewis, D. (1970). General semantics. *Synthese*, 22(1/2):18–67.
- Loffler, G. (2008). Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision research*, 48(20):2106–2127.
- Love, B. C., Rouder, J. N., and Wisniewski, E. J. (1999). A structural account of global and local processing. *Cognitive Psychology*, 38:291–316.
- Maloney, L. T. (2003). Surface colour perception and environmental constraints. In Mausfeld, R. and Heyer, D., editors, *Colour Perception: Mind and The Physical World*, chapter 9, pages 279–300. Oxford University Press, New York.
- Maloney, L. T. and Yang, J. N. (2003). The illuminant estimation hypothesis and surface color perception. In Mausfeld, R. and Heyer, D., editors, *Colour Perception: Mind and The Physical World*, chapter 11, pages 335–358. Oxford University Press, New York.
- Marr, D. (1976). Early processes of visual information. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 275(942):483–519.

- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press, Cambridge, MA.
- Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167):187–217.
- Marr, D. and Nishihara, H. K. (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, M. (2010). How things look (and what things look that way). In Nanay, B., editor, *Perceiving the World*, pages 226–253. Oxford University Press, New York.
- Mausfeld, R. (2003). Conjoint representations and the mental capacity for multiple simultaneous perspectives. In Hecht, H., Schwartz, R., and Atherton, M., editors, *Looking Into Pictures: An Interdisciplinary Approach to Pictorial Space*, pages 17–60. The MIT Press, Cambridge, MA.
- McCready, D. (1985). On size, distance, and visual angle perception. *Perception & Psychophysics*, 37(4):323–334.
- McLaughlin, B. P. (2009). Systematicity redux. *Synthese*, 170(2):251–274.
- McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology*, 18(1):87–121.
- Miller, G. A. and Isard, S. (1964). Free recall of self-embedded english sentences. *Information and Control*, 7:292–303.

- Montague, R. (1970). Universal grammar. *Theoria*, 36(3):373–398.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In Suppes, P., Moravcsik, J., and Hintikka, J., editors, *Approaches to Natural Languages*, pages 221–242. Dordrecht.
- Moore, C. M. and Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, 23(2):339–352.
- Müller, M., Malinowski, P., Gruber, T., and Hillyard, S. (2003). Sustained division of the attentional spotlight. *Nature*, 424:209–312.
- Naber, M., Carlson, T. A., Verstraten, F. A., and Einhäuser, W. (2011). Perceptual benefits of objecthood. *Journal of Vision*, 11(4):1–9.
- Nakayama, K., He, Z. J., and Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In Kosslyn, S. and Osherson, D., editors, *Invitation to Cognitive Science*, volume 2, pages 1–70. Cambridge, MA: MIT Press.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3):353–383.
- Navon, D. (1983). How many trees does it take to make a forest? *Perception*, 12(3):239–254.
- Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126.

- Nissen, M. J. (1985). Accessing features and objects: Is location special? In Posner, M. I. and Marin, O. S., editors, *Attention and Performance XI*, pages 205–219. Erlbaum, Hillsdale, NJ.
- Noë, A. (2004). *Action In Perception*. MIT Press, Cambridge, MA.
- Ochsner, K. N., Chiu, C. P., and Schachter, D. L. (1994). Varieties of priming. *Current Opinion in Neurobiology*, 4(2):189–194.
- Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9(4):441–474.
- Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In Roach, E. and Lloyd, B. B., editors, *Cognition and Categorization*, pages 259–303. Lawrence Elbaum Associates, Hillsdale, NJ.
- Palmer, S. E. (1980). What makes triangles point: Local and global effects in configurations of ambiguous triangles. *Cognitive Psychology*, 12(3):285 – 305.
- Palmer, S. E. (1983). The psychology of perceptual organization: A transformational approach. In Beck, J., Hope, B., and Rosenfeld, A., editors, *Human and Machine Vision*, pages 269–339. Academic Press.
- Palmer, S. E. (1989). Reference frames in the perception of shape and orientation. In Shepp, B. E., editor, *Object Perception: Structure & Process*, volume Ballesteros, Soledad, chapter 5, pages 121–163. Hillsdale, N.J.: Lawrence Erlbaum Associates.

- Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. The MIT Press, Cambridge, MA.
- Palmer, S. E. and Kimchi, R. (1986). The information processing approach to cognition. In Knapp, T. and Robertson, L., editors, *Approaches to Cognition: Contrasts and Controversies*, pages 37–77. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Partee, B. (1975). Montague grammar and transformational grammar. *Linguistic Inquiry*, 6(2):203–300.
- Partee, B. (2004). Compositionality. In *Compositionality in Formal Semantics*, pages 153–181. Blackwell.
- Peacocke, C. (1976). What is a logical constant? *The Journal of Philosophy*, 73(9):221–240.
- Peacocke, C. (1983). *Sense and Content: Experience, Thought, And Their Relations*. Oxford University Press, New York.
- Peacocke, C. (1992). *A Study of Concepts*. The MIT Press, Cambridge, MA.
- Peacocke, C. (1994). Content, computation, and externalism. *Mind & Language*, 9(3):303–335.
- Pelletier, F. J. (1994). On an argument against semantic compositionality. In *Logic and philosophy of science in Uppsala*, pages 599–610. Springer.
- Pelletier, F. J. (2012). Holism and compositionality. In Hinzen, W., Machery, E., and Werning,

- M., editors, *The Oxford Handbook of Compositionality*, pages 149–174. Oxford University Press.
- Pinker, S. (1980). Mental imagery and the third dimension. *Journal of Experimental Psychology: General*, 109(3):354–371.
- Pinker, S. (1988). A computational theory of the mental imagery medium. In Denis, M., Engelkamp, J., and Richardson, J. T., editors, *Cognitive and Neuropsychological Approaches to Mental Imagery*, pages 17–32. Martinus Nijhoff Publishers, Boston.
- Polat, U. and Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33(7):993–999.
- Pomerantz, J. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General*, 112(4):516–540.
- Pomerantz, J. R. and Kubovy, M. (1986). Theoretical approaches to perceptual organization. In Boff, K., Kaufman, L., and Thomas, J., editors, *Handbook of Perception and Human Performance*, volume 2, Cognitive Processes and Performance, pages 36–1 to 36–46. New York: Wiley.
- Potter, M. C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in Psychology*, 3:1–9.
- Prinz, J. J. (2007). The intermediate level theory of consciousness. In Velmans, M. and Schneider, S., editors, *The Blackwell Companion to Consciousness*, pages 248–260. Blackwell Publishing, Malden, MA.

- Prinzmetal, W. (1981). Principles of feature integration in visual perception. *Perception & Psychophysics*, 30(4):330–340.
- Putnam, H. (1967). The nature of mental states. In *Mind, Language and Reality: Philosophical Papers*, volume 2, pages 429–440. Cambridge University Press, New York.
- Pylyshyn, Z. (1979). Validating computational models: A critique of Anderson’s indeterminacy of representation claim. *Psychological Review*, 86(4):383–394.
- Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32(1):65–97.
- Pylyshyn, Z. (2003). *Seeing and Visualizing: It’s Not What You Think*. The MIT Press, Cambridge, MA.
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. The MIT Press, Cambridge, MA.
- Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25(2):157–182.
- Pylyshyn, Z. W. and Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3):179–197.
- Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, 129(5):643–673.

- Reiter, R. and Mackworth, A. K. (1989). A logical framework for depiction and image interpretation. *Artificial Intelligence*, 41(2):125–155.
- Rensink, R. A. and Enns, J. T. (1995). Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, 102(1):101–130.
- Rescorla, M. (2009). Cognitive maps and the language of thought. *British Journal for the Philosophy of Science*, 60(2):377–407.
- Rescorla, M. (2012). How to integrate representation into computational modeling, and why we should. *Journal of Cognitive Science*, 13:1–38.
- Rey, G. (1995). A not “merely empirical” argument for a language of thought. *Philosophical Perspectives*, 9, AI, Connectionism and Philosophical Psychology:201–222.
- Richards, W., Dawson, B., and Whittington, D. (1986). Encoding contour shape by curvature extrema. *Journal of the Optical Society of America A*, 3(9):1483–1491.
- Rock, I. (1973). *Orientation and Form*. Academic Press.
- Rock, I. (1985). *The Logic of Perception*. The MIT Press, Cambridge, MA.
- Rousselet, G. A., Fabre-Thorpe, M., and Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, 5:629–630.
- Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception*, 30(3):339–366.

- Schacter, D. L., Cooper, L. A., Delaney, S. M., Peterson, M. A., and Tharan, M. (1991). Implicit memory for possible and impossible objects: Constraints on the construction of structural descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1):3–19.
- Scharlau, I. (2004). Evidence for split foci of attention in a priming paradigm. *Perception & Psychophysics*, 66(6):988–1002.
- Schellenberg, S. (2008). The situation-dependency of perception. *The Journal of Philosophy*, 105(2):55–84.
- Schellenberg, S. (2018). *The Unity of Perception: Content, Consciousness, Evidence*. Oxford University Press, New York.
- Schier, E. (2007). The represented object of color experience. *Philosophical Psychology*, 20(1):1–27.
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80(1–2):1–46.
- Shapiro, S. (2005). Logical consequence, proof theory, and model theory. In Shapiro, S., editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 651–670. Oxford University Press, New York.
- Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and Brain Sciences*, 24(4):581–601.
- Shepard, R. N. and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703.

- Sher, G. (1996). Did Tarski commit “Tarski’s fallacy”? *The Journal of Symbolic Logic*, 61(2):653–686.
- Shimojima, A. (1996). Operational constraints in diagrammatic reasoning. In Allwein, G. and Barwise, J., editors, *Logical Reasoning with Diagrams*, pages 27–48. Oxford University Press, New York.
- Shimojima, A. (2001). The graphic-linguistic distinction. *Artificial Intelligence Review*, 15:5–27.
- Shimojo, S., Silverman, G. H., and Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision research*, 29(5):619–626.
- Shin, S.-J. (1994). *The Logical Status of Diagrams*. Cambridge University Press, New York.
- Singh, M. and Hoffman, D. D. (2001). Part-based representations of visual shape and implications for visual cognition. In Shipley, T. F. and Kellman, P. J., editors, *From fragments to objects: Segmentation and grouping in vision*, pages 401–459. New York: Elsevier Science.
- Skow, B. (2007). Are shapes intrinsic? *Philosophical Studies*, 133(1):111–130.
- Smith, A. D. (2002). *The Problem of Perception*. Harvard University Press, Cambridge, MA.
- Soames, S. (1987). Direct reference, propositional attitudes, and semantic content. *Philosophical Topics*, 15(1):47–87.
- Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14(1):29–56.

- Steinman, R. M., Pizlo, Z., and Pizlo, F. J. (2000). Phi is not beta, and why Wertheimer's discovery launched the Gestalt revolution. *Vision Research*, 40(17):2257–2264.
- Stevens, K. A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, 46(3):183–195.
- Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. The MIT Press, Cambridge, MA.
- Sugihara, K. (1986). *Machine Interpretation of Line Drawings*. Cambridge, MA: The MIT Press.
- Sutherland, N. S. (1968). Outlines of a theory of visual pattern recognition in animals and man. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 171(1024):297–317.
- Svenonius, P. (2010). Spatial P in English. In Cinque, G. and Rizzi, L., editors, *Mapping Spatial PPs: The Cartography of Syntactic Structures*, volume 6, pages 127–160. Oxford University Press, New York.
- Tacca, M. C. (2011). Commonalities between perception and cognition. *Frontiers in Psychology*, 2:1–10.
- Tarski, A. (1956). On the concept of logical consequence. In *Logic, Semantics, Metamathematics*, pages 409–420. Oxford University Press, Oxford.
- Tarski, A. and Corcoran, J. (1986). What are logical notions? *History and Philosophy of Logic*, 7(2):143–154.

- Thompson, B. J. (2010). The spatial content of experience. *Philosophy and Phenomenological Research*, 81(1):146–184.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Letters to Nature*, 381:520–522.
- Todd, J. T. (2004). The visual perception of 3d shape. *Trends in Cognitive Sciences*, 8(3):115–121.
- Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, 255(5):106–115.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology Section A*, 40(2):201–237.
- Treisman, A. and Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14:107–141.
- Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.
- Tsotsos, J. K. (1993). The role of computational complexity in perceptual theory. In Masin, S. C., editor, *Foundations of Perceptual Theory*, Advances in Psychology. Elsevier Science Publishers B.V., New York.
- Tulving, E. and Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247(4940):301–6.

- Tye, M. (1991). *The Imagery Debate*. The MIT Press, Cambridge, MA.
- Tye, M. (1996). Perceptual experience is a many-layered thing. *Philosophical Issues*, 7:117–126.
- Tye, M. (2005). In defense of representationalism: Reply to commentaries. In Aydede, M., editor, *Pain: New Essays on Its Nature and the Methodology of Its Study*, pages 163–175. The MIT Press, Cambridge, MA.
- Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405–426.
- Ullman, S. (1996). *High-Level Vision: Object Recognition and Visual Cognition*. The MIT Press, Cambridge, MA.
- van der Helm, P. A. and Leeuwenberg, E. L. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review*, 103(3):429–456.
- Varzi, A. C. (2002). On logical relativity. *Nous*, 36(s1):197–219.
- Vecera, S. (1994). Grouped locations and object-based attention: Comment on egly, driver, and rafal (1994). *Journal of Experimental Psychology: General*, 123(3):316–320.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., and von der Hedyt, R. (2012a). A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychological Bulletin*, 138(6):1172–1217.

- Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., and van Leeuwen, C. (2012b). A century of gestalt psychology in visual perception: Ii. conceptual and theoretical foundations. *Psychological Bulletin*, 138(6):1218–1252.
- Waltz, D. (1975). Understanding line drawings of scenes with shadows. In Winston, P., editor, *The Psychology of Computer Vision*, pages 19–91. McGraw-Hill, New York.
- Waltz, D. L. (1972). Generating semantic description from drawings of scenes with shadows. Technical report, MIT Artificial Intelligence Laboratory.
- Watt, R. (1988). *Visual Processing: Computational, Psychophysical, and Cognitive Research*. Hillsdale, NJ: Erlbaum.
- Wertheimer, M. (1938a). Gestalt theory. In Ellis, W. D., editor, *A Source Book of Gestalt Psychology*, pages 1–11. Routledge & Kegan Paul, London.
- Wertheimer, M. (1938b). Laws of organization in perceptual forms. In Ellis, W., editor, *A source book of Gestalt psychology*, pages 71–88. London: Routledge & Kegan Paul.
- Westerståhl, D. (1998). On mathematical proofs of the vacuity of compositionality. *Linguistics and Philosophy*, 21:635–643.
- Westheimer, G. (1981). Visual hypercauity. *Progress in Sensory Physiology*, 1:1–30.
- Williams, A. and Weisstein, N. (1978). Line segments are perceived better in a coherent context than alone: An object-line effect in visual perception. *Memory & Cognition*, 6(2):85–90.

- Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial intelligence*, 17(1):17–45.
- Wolfe, J. M. and Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37(1):25–43.
- Yuille, A. and Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7):301–308.
- Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and Philosophy*, 17(4):329–342.