

Published in final edited form as:

Psychol Bull. 2012 November ; 138(6): 1172–1217. doi:10.1037/a0029333.

A Century of Gestalt Psychology in Visual Perception I. Perceptual Grouping and Figure-Ground Organization

Johan Wagemans¹, James H. Elder², Michael Kubovy³, Stephen E. Palmer⁴, Mary A. Peterson⁵, Manish Singh⁶, and Rüdiger von der Heydt⁷

¹Laboratory of Experimental Psychology, University of Leuven (KU Leuven), Belgium and Institute of Advanced Studies (IEA-Paris), France

²Centre for Vision Research, York University, Canada

³Department of Psychology, University of Virginia, U.S.A.

⁴Department of Psychology, University of California at Berkeley, U.S.A.

⁵Department of Psychology and Cognitive Science Program, University of Arizona, U.S.A.

⁶Department of Psychology and Center for Cognitive Science, Rutgers University - New Brunswick, U.S.A.

⁷Mind/Brain Institute, Johns Hopkins University, U.S.A.

Abstract

In 1912, Max Wertheimer published his paper on phi motion, widely recognized as the start of Gestalt psychology. Because of its continued relevance in modern psychology, this centennial anniversary is an excellent opportunity to take stock of what Gestalt psychology has offered and how it has changed since its inception. We first introduce the key findings and ideas in the Berlin school of Gestalt psychology, and then briefly sketch its development, rise, and fall. Next, we discuss its empirical and conceptual problems, and indicate how they are addressed in contemporary research on perceptual grouping and figure-ground organization. In particular, we review the principles of grouping, both classical (e.g., proximity, similarity, common fate, good continuation, closure, symmetry, parallelism) and new (e.g., synchrony, common region, element and uniform connectedness), and their role in contour integration and completion. We then review classic and new image-based principles of figure-ground organization, how it is influenced by past experience and attention, and how it relates to shape and depth perception. After an integrated review of the neural mechanisms involved in contour grouping, border-ownership, and figure-ground perception, we conclude by evaluating what modern vision science has offered compared to traditional Gestalt psychology, whether we can speak of a Gestalt revival, and where the remaining limitations and challenges lie. A better integration of this research tradition with the rest of vision science requires further progress regarding the conceptual and theoretical foundations of the Gestalt approach, which will be the focus of a second review paper.

Keywords

Gestalt; grouping principles; figure-ground organization; neural mechanisms; vision science

1 Introduction

Exactly 100 years ago Wertheimer (1912) published his paper on phi motion—perception of pure motion, without object motion—which many consider to be the beginning of Gestalt psychology as an important school of thought. The present status of Gestalt psychology is ambiguous. On the one hand, many psychologists believe that the Gestalt school died with its founding fathers in the 1940s or after some devastating empirical findings regarding electrical field theory in the 1950s, or that it declined because of fundamental limitations that blocked further progress, while stronger theoretical and experimental frameworks arose in the 1960s and 1970s that have dominated the field ever since (e.g., cognitive science, neuroscience). On the other hand, almost all psychology textbooks still contain a Gestalt-like chapter on perceptual organization (although often poorly connected to the other chapters), and new empirical papers on Gestalt phenomena appear with increasing frequency.

We are convinced that Gestalt psychology is still relevant to current psychology in several ways. First, questions regarding the emergence of structure in perceptual experience and the subjective nature of phenomenal awareness (e.g., visual illusions, perceptual switching, context effects) continue to inspire contemporary scientific research, using methods and tools that were not at the Gestaltists' disposal. Second, the revolutionary ideas of the Gestalt movement continue to challenge some of the fundamental assumptions of mainstream vision science and cognitive neuroscience (e.g., elementary building blocks, channels, modules, information-processing stages). Much progress has been made in the field of nonlinear dynamical systems, both theoretically and empirically (e.g., techniques to measure and analyze cortical dynamics), progress that allows modern vision scientists to surpass some of the limitations in old-school Gestalt psychology as well as those in mainstream vision research.

The centennial anniversary of Gestalt psychology is therefore an excellent opportunity to take stock of what we have discovered about core Gestalt phenomena of perceptual organization and how our understanding of the underlying mechanisms has evolved since Wertheimer's seminal contribution. Due to this review's scope, we divide it in two parts: This paper deals with perceptual grouping and figure-ground organization, whereas the second covers modern developments regarding the general conceptual and theoretical frameworks that underlie Gestalt ideas (e.g., holism, emergence, dynamics, simplicity). In Table 1, we provide an overview of the topics covered in the first review paper, together with the section headings, the questions or issues being raised, and some of the answers provided. One of the aims of our review is to remove the many misunderstandings surrounding Gestalt psychology, which are listed separately in Table 2, along with a more balanced view on the actual state of affairs.

To put these two reviews in perspective, we first introduce the key findings and ideas of the founders of Gestalt psychology, along with a brief sketch of its further development, rise, and fall (for an extensive treatment, see Ash, 1995). The historical section ends with a discussion of Gestalt psychology's empirical and conceptual problems and an indication of how these limitations are being addressed in current research. We then review current research on perceptual grouping and figure-ground organization in more detail in the remaining sections. We focus on these two topics because they were the most important ones in the Gestalt tradition of perceptual organization and still are today, even for vision in general.

2 A Brief History of Gestalt Psychology

This section addresses four questions regarding Gestalt psychology: (1) How did it start? (2) What does it stand for? (3) How did it evolve? (4) Where does it stand now?

2.1 The Emergence of Gestalt Psychology

What Max Wertheimer discovered in 1912 was called phi motion, a special case of apparent motion. (For an excellent discussion of its historical importance, see Sekuler, 1996; for a demonstration of the phenomenon and for a review of its misrepresentation in later sources, see Steinman, Pizlo, & Pizlo, 2000.) According to the conventional view of apparent motion, we see an object at several successive positions and motion is then “added” subjectively. If this were correct, then an object would have to be seen as moving, and at least two positions—the starting and end points—would be required to produce seen motion. Neither of these conditions held in the case of phi motion. In the key experiment, a white strip was placed on a dark background in each of two slits in the wheel of a tachistoscope, and the rotation speed was adjusted to vary the time required for the light to pass from one slit to the next (i.e., the interval between the two). Above a certain threshold value (~200 ms), observers saw the two lines in succession. With much shorter intervals (~30 ms), the two lines appeared to flash simultaneously. At the optimal stage (~60 ms), observers perceived a motion that could not be distinguished from real motion. When the interval was decreased slightly below 60 ms, after repeated exposures, observers perceived motion without a moving object—that is, pure phenomenal or phi motion. Although only three observers were tested, “the characteristic phenomena appeared in every case unequivocally, spontaneously, and compellingly” (Wertheimer 1912/1961, p. 1042). In the same paper, Wertheimer proposed a physiological model described in terms of a short circuit and a flooding back of the current flow (“transverse functions of a special kind;” Wertheimer, 1912/1961, p. 1085), which produced what he called “a unitary continuous whole-process” (Wertheimer, 1912/1961, p. 1087). He then extended this theory to the psychology of pure simultaneity (for the perception of form or shape) and of pure succession (for the perception of rhythm or melody). These extensions were decisive for the emergence of Gestalt theory.

2.2 Essentials of Gestalt Theory

The phi phenomenon was the perception of a pure process, a transition that could not be composed from more primitive percepts of a single object at two locations. In other words, perceived motion was not added subjectively after the sensory registration of two spatiotemporal events but had its own phenomenological characteristics and ontological status. From this phenomenon, Wertheimer concluded that structured wholes or Gestalten, rather than sensations, are the primary units of mental life. This was the key idea of the new and revolutionary Gestalt theory, developed by Wertheimer and his colleagues in Berlin. An overview of how the Berlin school of Gestalt psychology distinguished itself from the dominant view of structuralism and empiricism, as well as of related Gestalt schools is given in Table 3.

The notion of “Gestalt” had already been introduced into psychology by Christian von Ehrenfels in his essay “On Gestalt qualities” (1890). Based on the observation that humans can recognize two melodies as identical even when no two corresponding notes in them have the same frequency, von Ehrenfels argued that these forms must possess a “Gestalt quality”—a characteristic that is immediately given, along with the elementary sensations that serve as its foundation, a characteristic that is dependent on its constituent objects but rises above them. For von Ehrenfels, Gestalt qualities rest uni-directionally on sense data: Wholes are more than the sums of their parts, but the parts are the foundation (“Grundlage”) of the whole. In contrast, Wertheimer claimed that functional relations determine what will

appear as the whole and what will appear as parts (i.e., reciprocal dependency). Often the whole is grasped even before the individual parts enter consciousness. The contents of our awareness are by and large not additive but possess a characteristic coherence. They are structures that are segregated from the background, often with an inner center, to which the other parts are related hierarchically. Such structures or “Gestalten” are different from the sum of the parts. They arise from continuous global processes in the brain, rather than combinations of elementary excitations.

With this step, Wertheimer separated himself from the Graz school of Gestalt psychology, represented by Alexius Meinong, Christian von Ehrenfels, and Vittorio Benussi. They maintained a distinction between sensation and perception, the latter produced on the basis of the former. The Berlin school, represented by Max Wertheimer, Kurt Koffka, and Wolfgang Köhler, considered a Gestalt as a whole in itself, not founded on any more elementary objects. In their view, perception was not the product of sensations but it arose through dynamic physical processes in the brain. As a result, the Berlin school also rejected stage theories of perception proposed by the Leipzig school, represented by Felix Krüger and Friedrich Sander, in which the gradual emergence of Gestalten (“Aktualgenese” or “microgenesis”) played a central role. Although the Berlin theorists adhered to a nonmechanistic theory of causation and did not analyze the processes into stages, they did believe that the functional relations in the emergence of Gestalts could be specified by laws of perceptual organization.

2.3 Further Development, Rise, and Fall of Gestalt Psychology

Two major developments are generally considered as highlights in the history of Gestalt psychology: Köhler’s discussion of “physical Gestalten” (1920) and Wertheimer’s proposal of “Gestalt laws of perceptual organization” (1923). Köhler (1920) extended the Gestalt concept from perception and behavior to the physical world, thus attempting to unify holism (i.e., the doctrine stressing the importance of the whole) and natural science. He proposed to treat the neurophysiological processes underlying Gestalt phenomena in terms of the physics of field continua rather than that of particles or point-masses. In such continuous field systems, which he called strong Gestalten, the mutual dependence among the parts is so great that no displacement or change of state can occur without influencing all the other parts of the system. Köhler showed that stationary electric currents, heat currents, and all phenomena of flow are strong Gestalten in this sense. These he distinguished from what he called weak Gestalten, which do not show this mutual interdependence.

In addition, Köhler tried to construct a specific testable theory of brain processes that could account for perceived Gestalten in vision. He thought of visual Gestalten as the result of an integrated process in what he referred to as “the entire optical sector,” including retina, optical tract, and cortical areas, as well as transverse functional connections among conducting nerve fibers (i.e., a recurrent neural network in modern terms). He proposed an electrical field theory, in which “the lines of flow are free to follow different paths within the homogeneous conducting system, and the place where a given line of flow will end in the central field is determined in every case by the conditions in the system as a whole” (Köhler, 1920/1938, p. 50). In modern terms, Köhler had described the visual system as a self-organizing physical system.

These ideas led Köhler to postulate a psychophysical isomorphism between the psychological reality and the brain events underlying it: “actual consciousness resembles in each case the real structural properties of the corresponding psycho-physiological process” (Köhler, 1920/1938, p. 38). By this he meant functional instead of geometrical similarity indicating that brain processes do not take the form of the perceived objects themselves. In addition, he insisted that such a view does not prescribe complete homogeneity of the cortex

but is perfectly compatible with functional articulation. Experiments to establish the postulated connections between experienced and physical Gestalten in the brain were at the time nearly impossible to conduct, but decades later, Köhler attempted to do so (see below).

Around the same time, Max Wertheimer further developed his Gestalt epistemology and outlined the research practice of experimental phenomenology that was based on it. He first stated the principles publically in a manifesto published in Volume 1 of *Psychologische Forschung* in 1922. Wertheimer called for descriptions of conscious experience in terms of the units people naturally perceive, rather than the artificial ones imposed by standard scientific methods. By assuming that conscious experience is composed of units analogous to physical point-masses or chemical elements, psychologists constrain themselves to a piecemeal inquiry into the contents of consciousness, building up higher entities from constituent elements, using associative connections. In fact, such and-summations (“Und-Summe”), as Wertheimer called them, appear “only rarely, only under certain characteristic conditions, only within very narrow limits, and perhaps never more than approximately” (Wertheimer 1922/1938, p. 13). Rather, what is given in experience “is itself in varying degrees ‘structured’ (‘gestaltet’), it consists of more or less definitely structured wholes and whole-processes with their whole-properties and laws, characteristic whole-tendencies and whole-determinations of parts” (Wertheimer 1922/1938, p. 14). The perceptual field does not appear to us as a collection of disjointed sensations, but possesses a particular organization of spontaneously combined and segregated objects.

In 1923, Wertheimer published a follow-up paper, which was an attempt to elucidate the fundamental principles of that organization. The most general principle was the so-called law of *Prägnanz*, stating, in its most general sense, that the perceptual field and objects within it will take on the simplest and most encompassing (“ausgezeichnet”) structure permitted by the given conditions. For Köhler (1920), this tendency towards the *Prägnanz* of the Gestalt was just another example that phenomenal Gestalten were like physical Gestalten: As shown by Maxwell and Planck, all processes in physical systems, left to themselves, show a tendency to achieve the maximal level of stability (homogeneity, simplicity, symmetry) with the minimum expenditure of energy allowed by the prevailing conditions. More specific principles that determine perceptual organization according to Wertheimer were proximity, similarity, uniform density, common fate, direction, good continuation and “whole properties” (or “*Ganzeigenschaften*”) such as closure, equilibrium, and symmetry.

Empirical work on these principles existed before Wertheimer’s landmark paper (for a recent review, see Vezzani, Marino, & Giora, 2012), but now the general claim that perceptual experience is organized was turned into a complex open-ended research program aimed at the discovery of the laws or principles governing perceptual organization in both its static and dynamic aspects. It is this research program that Wertheimer, Koffka and Köhler started to work on with their students, once they had acquired professorships at major universities in Germany in the 1920s and 1930s. We cannot cover this flourishing period of the Berlin school of Gestalt psychology extensively here, but a few highlights that deserve mentioning in passing are studies by Kurt Gottschaldt on embedded figures (1926), Joseph Ternus on phenomenal identity (1926), Karl Duncker on induced motion (1929), Wolfgang Metzger on a homogeneous *Ganzfeld* (1930) and motion in depth (1934). In the meantime, Gestalt thinking also affected research on other sense modalities (e.g., binaural hearing by Erich von Hornbostel), on learning and memory (e.g., Otto von Lauenstein and Hedwig von Restorff), and on thought (e.g., Karl Duncker). Later, Gestalt theory was also applied to action and emotion (by Kurt Lewin), to neuropathology and the organism as a whole (by Adhemar Gelb and Kurt Goldstein), and to film theory and aesthetics (by Rudolf Arnheim).

This period marked the high point but not the end of Gestalt psychology's theoretical development, its research productivity, and its impact on German science and culture.

Around this time, Gestalt theory also started to have some impact on research in the U.S., mainly owing to Wolfgang Köhler and Kurt Koffka (see King & Wertheimer, 2005, Chapter 10). For instance, Koffka's (1935) notion of vector fields inspired some interesting empirical work published in the *American Journal of Psychology* (Brown & Voth, 1937; Orbison, 1939). Reviews of Gestalt psychology appeared in *Psychological Review* on a regular basis (e.g., Helson, 1933; Hsiao, 1928), a comprehensive book on state-of-the-art Gestalt psychology was published as early as 1935 (Hartmann, 1935), and three years later Ellis's (1938) influential collection of translated excerpts of core Gestalt readings made some of the original sources accessible to a non-German-speaking audience. Already in 1922, at Robert Ogden's invitation, Koffka had published a full account of the Gestalt view on perception in *Psychological Bulletin*.

At first sight, Gestalt theory seemed to develop rather consistently, from studying the fundamental laws of psychology first under the simplest conditions, in elementary problems of perception, before including complex sets of conditions, and turning to other domains such as memory, thinking, emotion, aesthetics, and so forth. At the same time, however, the findings obtained did not always fit the original theories, which posed serious challenges to the Gestalt framework. Even more devastating to the development of Gestalt psychology was the emergence of the Nazi regime in Germany from 1933 to World War II. In this period, many of the psychology professors at German universities lost their posts because of the discrimination and prosecution of Jews, so they emigrated to the U.S. to take on new positions there. The works by German psychologists who stayed, for instance, Edwin Rausch's monograph on "summative" and "nonsummative" concepts (1937) and Wolfgang Metzger's (1941) psychology textbook, were largely ignored outside Germany. Metzger's synoptic account of research on the Gestalt theory of perception entitled "Gesetze des Sehens" ("Laws of seeing"), first published in 1936 and later reissued and vastly expanded three times, was only translated into English in 2006.

After emigrating to the U.S., the founding fathers of Gestalt psychology did not perform many new experiments. Instead, they mainly wrote books in which they outlined their views (e.g., Koffka, 1935; Köhler, 1940; Wertheimer, 1945). The major exception was Köhler who had taken up physiological psychology using EEG recording and other methods in an attempt to directly verify his isomorphism postulate. Initially, his work with Hans Wallach on figural aftereffects appeared to support his interpretation in terms of satiation of cortical currents (Köhler & Wallach, 1944). Afterwards, he was able to directly measure cortical currents—as EEG responses picked up from electrodes at the scalp—whose flow direction corresponded to the direction of movement of objects in the visual field (Köhler & Held, 1949).

Soon after that breakthrough, however, Lashley and colleagues (Lashley, Chow, & Semmes, 1951) performed a more critical test of Köhler's electric field theory and its underlying postulate of isomorphism. If the flows of current picked up from the scalp in Köhler and Held's experiments indeed reflected the organized pattern of perception and not merely the applied stimulation, and if that pattern of perception would result from a global figure-field across the whole cortex, a marked alteration of the currents should distort the perception of these visual figures. By inserting metallic strips and metal pins in large regions of the visual cortex of rhesus monkeys, Lashley et al. could short-circuit the cortical currents. Surprisingly, the monkeys could still perform the learned shape discriminations, demonstrating that global cortical currents were not a necessary condition for pattern perception. In subsequent experiments, Sperry and colleagues (Sperry, Miner, & Myers,

1955) performed extensive slicing and dense impregnation with metallic wires across the entire visual cortex of cats, and showed that these animals too could still perform rather difficult shape discriminations (e.g., between a prototypical triangle and distorted variants). Together, these two studies effectively ruled out electrical field theory as an explanation of cortical integration and undermined the empirical basis of any isomorphism between cortical flows of current and organized patterns of perception. Köhler (1965) naturally reacted to these developments but his counterarguments and suggestions for further experiments were largely ignored, and to most scientists at the time, the matter was closed. Electrical field theory, which had been one of the pillars of Gestalt psychology's scientific basis, was considered dead and buried.

While Gestalt psychology declined in the English-speaking world after World-War II, Italy remained a stronghold of Gestalt psychology. For instance, Metzger dedicated the third edition of his "Gesetze des Sehens" to his "Italian and Japanese friends." Among his friends were Musatti, Metelli, and Kanizsa—three major figures in Italian psychology. In spite of being Benussi's student and successor (from the Graz school), Cesare Musatti was responsible for introducing the Berlin school of Gestalt psychology in Italy and training important students in this tradition—most notably Metelli and Kanizsa, whose contributions continue to be felt today. Fabio Metelli is best known for his work on the perception of transparency (e.g., Metelli, 1974). Gaetano Kanizsa's most famous studies were performed in the 1950s with papers on subjective contours (e.g., the so-called Kanizsa triangle), modes of color appearance, and phenomenal transparency (Kanizsa, 1954, 1955a, 1955b), although their impact came much later, when he started to publish in English (Kanizsa, 1976, 1979).

In addition to Italy, Gestalt psychology was also strong in Belgium and in Japan. Albert Michotte became famous for his work on the perception of causality (1946/1963), arguing strongly against an inferential, associationist, empiricist account of it, like other Gestalt psychologists had done for other aspects of perception. For him, causality is perceived directly, not derived from more primitive sensations through some cognitive operation, and this percept could be shown to be tightly coupled to specific higher-order attributes in the spatiotemporal events presented to observers. He also introduced the notions of modal and amodal completion (Michotte, Thinès, & Crabbé, 1964), and studied several configural influences on these processes. (For a further discussion of Michotte's heritage, see Wagemans, van Lier, & Scholl, 2006.) Building on earlier collaborations of Japanese students with major German Gestalt psychologists (e.g., Sakuma with Lewin, Morinaga with Metzger), Gestalt psychology continued to develop in Japan after World-War II. For instance, Oyama did significant work on figural aftereffects (e.g., Sagara & Oyama, 1957) and perceptual grouping (e.g., Oyama, 1961).

2.4 The Current Status of Gestalt Psychology

Despite signs of well-deserved respect in the U.S. and in Germany (e.g., Köhler's APA presidency in 1957; Wertheimer's posthumous Wilhelm Wundt Medal in 1983), the ideas of the Gestaltists were received with ambivalence. On the one hand, they were recognized for raising central issues and provoking important debates in psychology, theoretical biology, and other fields, but on the other hand, their mode of thinking and research style did not sit comfortably in the intellectual and social climate of the postwar world, and they were confronted with vehement criticism. Two sets of explanations have been given for this outcome (Ash, 1995). The first emphasizes institutional, political, and biographical contingencies. Koffka, Köhler and Wertheimer all left for the U.S. and obtained positions where they could do excellent research but could not train PhDs. The Gestalt school's further expansion was also handicapped by the early deaths of Max Wertheimer in 1943 and Kurt Koffka in 1941, as well as many other Gestalt psychologists of the first and second generation (e.g., Duncker, Gelb, Lauenstein, Lewin, von Restorff). In Germany, Metzger,

Rausch, and Gottschaldt did have a large number of PhD students, but few of them carried on in the Gestalt tradition. A notable exception is Lothar Spillmann, who obtained his D. Phil. with Metzger in Münster in 1964 and who pioneered the impact of Gestalt ideas in modern neurophysiology ever since (e.g., Spillmann, 1999, 2009).

The second set of explanations concerns scientific issues of a methodological and conceptual nature (summarized in the left column of Table 4). Compared to the rigor of psychophysics and behaviorism, Gestalt psychology was severely criticized for offering mere demonstrations, using either very simple or confounded stimuli, formulating laws with little precision, and adding new “laws” for every factor shown to have an influence on perceptual organization. In the 1950s and 1960s, its critics increasingly insisted on causal explanations, by which they meant cognitive operations in the mind that could be modeled as computer algorithms or neural mechanisms that could be attributed to the properties of single cells that were discovered by Hubel and Wiesel in that period. In addition, serious conceptual limitations appeared when Gestalt thinking was extended to other areas such as personality and social psychology (e.g., Richard Crutchfield, Solomon Asch, Fritz Heider, David Krech). The further the metaphors were stretched, the harder it became to connect them to Köhler’s concept of a self-organizing brain and his speculations about electromagnetic brain fields.

Despite these criticisms, Gestalt thinking did not disappear from the stage completely. In the slipstream of Shannon’s information theory, a few researchers tried to provide a quantitative underpinning to the central Gestalt notion of simplicity (e.g., Attneave, 1954; Attneave & Arnoult, 1956; Hochberg & McAlister, 1953; Leeuwenberg, 1969, 1971; for a review, see Hatfield & Epstein, 1985). A number of independent, original scientists working on perception and information processing kept some Gestalt issues on the research agenda (e.g., Fred Attneave, Wendell Garner, Julian Hochberg, Irvin Rock). These became more prominent again with the discovery of true Gestalt phenomena such as global precedence in hierarchical letters (e.g., Navon, 1977), configural superiority effects based on emergent features (e.g., Pomerantz, Sager, & Stoever, 1977), and the importance of hierarchical structure in perceptual representations (e.g., Palmer, 1977). The experimental paradigms were derived from standard methods in cognitive psychology, and the results were incorporated into mainstream information-processing accounts (e.g., Beck, 1982; Kubovy & Pomerantz, 1981). In the major alternative approaches to visual perception—the ecological (e.g., Gibson, 1971) and computational (e.g., Marr, 1982) approaches—the influence of Gestalt thinking has also been acknowledged explicitly. In the last two or three decades, perceptual grouping and figure-ground organization—the most central topics of Berlin school research—have returned to center stage (e.g., Kimchi, Behrmann, & Olson, 2003), although the relationship to the original Gestalt theory (e.g., two-sided dependency between wholes and parts, minimum principle) is not always clear.

In the remainder of this paper, as well as in a second more theoretically oriented paper (Wagemans et al., 2012), we review the later developments in more detail (summarized in the right column of Table 4). We start with research on perceptual grouping in simple displays (Section 3) and extend this to contour grouping, integration, and completion in more complex shapes and real-world images (Section 4). In the next section, we cover research on figure-ground perception, where many of the factors affecting grouping, in addition to unique factors, exert an influence (Section 5). Although links to neural mechanisms are mentioned throughout, we also provide a more integrated account of the literature on the neural mechanisms of contour grouping and figure-ground organization in a separate section (Section 6). This review demonstrates that research from the last two or three decades has addressed (and partially solved) some of the major methodological and conceptual shortcomings in old-school Gestalt psychology.

3 Perceptual Grouping

3.1 Introduction

Historically, the visual phenomenon most closely associated with perceptual organization is grouping: the fact that observers perceive some elements of the visual field as “going together” more strongly than others. Indeed, perceptual grouping and perceptual organization are sometimes presented as though they were synonymous. They are not. Grouping is one particular kind of organizational phenomenon, albeit a very important one. Another is figure-ground organization. In general, grouping determines what the qualitative elements of perception are, and figure-ground determines the interpretation of those elements in terms of their shapes and relative locations in the layout of surfaces in the 3-D world.

Max Wertheimer first posed the problem of perceptual grouping in his ground-breaking 1923 paper by asking what stimulus factors influence the perceived grouping of discrete elements. He first demonstrated that equally spaced dots do not group together into larger perceptual units, except as a uniform line (Figure 1A) and then noted that when he altered the spacing between adjacent dots so that some dots were closer than others, the closer ones grouped together strongly into pairs (Figure 1B). This factor of relative distance, which Wertheimer called *proximity*, was the first of his famous laws or (more accurately) principles of grouping.

Wertheimer went on to illustrate other grouping principles, several of which are portrayed in Figure 1. Parts C, D, and E demonstrate different versions of the general principle of *similarity*: all else being equal, the most similar elements (in color, size, and orientation for these examples) tend to be grouped together. Another powerful grouping factor is *common fate*: all else being equal, elements that move in the same way tend to be grouped together. Notice that both common fate and proximity can actually be considered special cases of similarity grouping, with velocity and position as the relevant properties, respectively. Further factors influencing perceptual grouping of more complex elements, such as lines and curves, include *symmetry* (Figure 1G), *parallelism* (Figure 1H), and *continuity* or *good continuation* (Figure 1I). Continuity is important in Figure 1I because observers perceive it as containing two continuous intersecting lines rather than as two angles whose vertices meet at a point. Figure 1J illustrates the effect of *closure*: all else being equal, elements that form a closed figure tend to be grouped together. This display also shows that closure can dominate continuity, since the very same elements that were organized as two intersecting lines in Figure 1I are now organized as two angles meeting at a point in part Figure 1J.

One might think that such grouping principles are mere textbook curiosities only distantly related to normal perception. On the contrary, they pervade virtually all perceptual experiences because they determine the objects and parts that people perceive in the environment. (Hence, they also affect other sensory modalities; for a thorough discussion of grouping principles in audition, see Bregman, 1990; for a recent review of Gestalt principles in tactile perception, see Galace & Spence, 2011.) A practical application of the Gestalt principles is camouflage, which results when the same grouping processes that would normally make an organism stand out from its environment as a separate object, cause it to be grouped together with its surroundings instead. For instance, the same leopard that is clearly visible when it is seen in a tree against the uniform sky is difficult to see against a mottled, leafy backdrop—until it moves. Even perfect static camouflage is undone by the principle of common fate. In sum, camouflage and camouflage breaking provide an ecological rationale for the principles of grouping.

Since the early days of Gestalt psychology, considerable progress has been made, including (1) the discovery of additional principles, (2) the experimental measurement of the strength of grouping factors and the development of quantitative laws, as well as (3) new insights into the level of processing at which perceptual grouping happens. These new developments will be described in the next section, where we will also discuss their possible structural and ecological basis.

3.2 New Principles of Grouping

3.2.1 Generalized common fate—One of the most powerful of the classic grouping principles is common fate—the tendency for elements that “move together” to be perceived as a unitary entity (Wertheimer, 1923). The possibility that Wertheimer may have had a much broader range of phenomena in mind, however, is suggested by a passage of his seminal article that is not widely known because it was not included in Ellis’s translation (Wertheimer, 1923/1938): “Also this principle [of common fate] is valid in a wide range of conditions; how wide is not yet investigated here” (Wertheimer, 1923, p. 316; our own translation). In this vein, Sekuler and Bennett (2001) presented an extension of common fate to grouping by common luminance changes. They found that when elements of a visual scene become brighter or darker simultaneously, even if they have different luminances throughout, observers have a powerful tendency to group those elements perceptually. It is as though the principle of common fate operates not only for the common motion of elements through 3-D physical space, but through luminance space as well. The structural rationale for generalized common fate is clear: It is another example of similarity grouping, but based on similarity of *changes* in feature values, such as luminance or position, rather than on the similarity of the feature values themselves. An ecological rationale for grouping by common luminance changes might lie in the simultaneous brightening or darkening that occurs across a spatial area when the level of illumination changes (e.g., with the appearance of sunlight or shadows; see also van den Berg, Kubovy, & Schirillo, 2011).

3.2.2 Synchrony—Synchrony is the tendency for elements that change simultaneously to be grouped together (Alais, Blake & Lee, 1998; Lee & Blake, 1999). The changes do not have to be in the same direction, however, as they do in generalized common fate. A random field of black and white dots whose luminances change in polarity randomly over time against a gray background, for example, will segregate into two distinct regions if the dots in one area change synchronously rather than randomly. Grouping by synchrony can be considered as an even more general form of common fate in which the simultaneous changes do not have to involve either motion, as in classic common fate, or common direction of change, as in generalized common fate. The structural basis for grouping by synchrony is clear: the simultaneous occurrence of visible changes of the elements that are grouped. Such grouping makes sense because it reflects a strong temporal regularity in the stimulus event.

The ecological rationale behind grouping by synchrony is far less clear, however. Objects in the natural environment seldom change their properties in different directions or along different dimensions in temporal synchrony. Indeed, it is difficult even to devise plausible examples of ecological situations that would exhibit this kind of temporal regularity without some form of extended common fate being involved. Nevertheless, synchrony grouping may arise from some very general nonaccidentalness detection mechanism, possibly connected to the perception of causality (e.g., Michotte, 1946/1963). The argument is that the temporal coincidence of multiple changes is unlikely to be due to chance alone, and so it must have some common underlying cause related to an ecological event that relates the synchronously changing elements.

A radically different and quite controversial, rationale is that temporal synchrony of changes drives grouping because synchrony of neural firings is the physiological mechanism by which the brain codes all forms of grouping (e.g., Milner, 1974; von der Malsburg, 1981). The argument is that if the environment drives the neural substrate to produce synchronous firing by virtue of synchronous changes, the changing elements will automatically be grouped because of the synchronous firing. Some researchers report evidence that seems to support this claim (e.g., Gray & Singer, 1989; Singer & Gray, 1995), but others disagree (e.g., Shadlen & Movshon, 1999). This issue is discussed further in the second paper (Wagemans et al., 2012; Section 4). Further controversy surrounds synchrony grouping because it has been claimed that such grouping effects are actually produced by stimulus artifacts that can be detected by the early visual system (Farid, 2002; Farid & Adelson, 2001). These challenges are complex, but the bottom line is that both the existence of grouping synchrony and the mechanism by which it occurs are currently unclear. In general, these controversies show quite clearly that the interest in perceptual grouping principles remains strong in contemporary research.

3.2.3 Common region—Common region is the tendency for elements that lie within the same bounded area (or region) to be grouped together (Palmer, 1992). An illustration is provided in Figure 1K, where the black dots that lie within the same ovals are likely to be grouped into pairs. The structural basis for grouping by common region appears to be that all the elements within a given region share the topological property of being “inside of” or “contained by” some larger surrounding contour. If it is viewed as similarity of containment, it can be related to several other grouping principles based on similarity (e.g., color, orientation, and size). Common region also appears to have an ecological rationale arising from textures and hierarchically embedded parts. When a bounded region encloses a number of image elements, they are likely to be elements on the surface of a single object, such as a leopard’s spots or the features of a face, rather than independent objects that just happen accidentally to lie within the same bounding contour.

Experimental evidence for the existence of common region as a grouping factor comes from studies using the Repetition Discrimination Time or RDT method (Beck & Palmer, 2002; Palmer & Beck, 2007). In a speeded discrimination task, observers were able to report the shape of a repeated element more quickly in a line of otherwise alternating shapes (e.g., squares and circles) when the repeated shapes were located within the same surrounding region than when they were located in two separate regions.

3.2.4 Element connectedness—Element connectedness is the tendency for distinct elements that share a common border to be grouped together. The important structural basis for this form of grouping is the topological property of connectedness (Palmer & Rock, 1994). Connectedness can be considered as the limiting case of the classic factor of proximity, but Palmer and Rock argued that framing it this way puts the cart before the horse, in the sense that one needs distinct units to speak meaningfully about their distance in the first place. The compelling rationale for element connectedness is ecological: Pieces of matter that are physically connected to each other in 3-D space are the primary candidates for being parts of the same object, largely because they tend to behave as a single unit. The bristles, metal band, and handle of a paint brush, for example, constitute a single object in large part because of their connectedness, as demonstrated by the fact that when you push one part, the other parts move rigidly along with it.

The effectiveness of element connectedness was demonstrated in a behavioral task using the RDT method (Palmer & Beck, 2007). As was the case for grouping by common region, displays with elements that were connected to each other produced reliably faster responses than displays with unconnected elements. Another behavioral result that provides striking

support for the importance of element connectedness comes from a neuropsychological study by Humphreys and Riddoch (1993). Their patient, who was afflicted with Balint's syndrome—a condition resulting from bilateral damage to parietal cortex that results in a deficit in perceiving more than a single “object” at any given time—was unable to discriminate between arrays containing many circles of just one color (either all red or all green) and arrays in which half of the circles were red and the other half green. However, if pairs consisting of one red circle and one green circle were connected by lines, the same patient was able to make the discrimination between one-color and two-color arrays. Unifying a pair of circles through element connectedness thus appears to enable these patients to perceive them as a single perceptual object so that they could see two circles at once, a feat that was impossible for them in the absence of the connecting line.

3.2.5 Uniform connectedness—This principle represents something of a departure from standard Gestalt ideas about perceptual organization because it addresses the question of how the initial organization into elements might occur. In his classic article on grouping, Wertheimer (1923) never actually mentioned where the to-be-grouped elements came from. Presumably, he believed that they were somehow derived from the grouping principles he articulated, but Palmer and Rock (1994) argued that they arise from the earlier organizational process of uniform connectedness (UC), which is the principle by which the visual system initially partitions an image into a set of mutually exclusive connected regions having uniform (or smoothly changing) properties, such as luminance, color, texture, motion, and depth. The UC elements thus created form the entry level units into a part-whole hierarchy that is created by grouping together different UC regions and, if necessary, by parsing them into lower-level elements at deep concavities (e.g., Hoffman & Richards, 1984).

Palmer and Rock's claims regarding the foundational status of UC have not been uniformly accepted. Peterson (1994), for instance, argued that UC is one of many properties relevant to partitioning the visual field, and that UC units are not entry-level units. Kimchi (2000) examined the role of UC in experiments designed to reveal the gradual emergence, or microgenesis, of organizational processes using a primed matching task with displays containing connected or disconnected elements. The complex results she obtained were not consistent with UC being the sole determinant of entry-level units in a part-whole hierarchy, as Palmer and Rock (1994) proposed. Rather, they showed that collinearity and closure were at least as important, if not more so, in the initial organization that can be tapped by such methods. Nevertheless, the theoretical rationale for some organizational process like UC to contribute to creating a set of potential perceptual units on which further grouping and parsing can operate seems sound. Indeed, something like it is a standard assumption in most theories of computational vision (e.g., Marr, 1982).

Whereas the research on grouping principles, as described above, provides predictions about what elements in a display are likely to be grouped together, it does not reveal how strong each of the grouping principles is. This is the focus of the next sections, covering studies applying static and dynamic stimuli, respectively.

3.3 Grouping Principles in Discrete Static Patterns

3.3.1 Conceptual background—Starting with Wertheimer (1923), researchers have often used *dot lattices* to quantify grouping. A dot lattice is a collection of dots in the plane which is invariant under at least two translations, a (with length $|a|$) and b (whose length is $|b| \geq |a|$). These two lengths, and the angle between the vectors, γ (constrained by $60^\circ \leq \gamma \leq 90^\circ$), define the basic structure of the lattice, by defining the parallelogram between each quartet of dots in the lattice (Kubovy, 1994, extending the work of Bravais, 1850/1949; see

Figure 2A). The diagonals of this parallelogram are denoted c and d (where $|c| \leq |d|$). In its canonical orientation, a is horizontal. More generally, the orientation of the lattice can be defined by the angle θ (measured counterclockwise) and $|a|$ is called the scale of the lattice. Since scale is irrelevant to the invariant properties of the lattice and unimportant for grouping over a reasonable range, the relevant parameters are $|b|/|a|$ and γ (Figure 2B). With these two parameters, six different types of lattices can be defined, each characterized by their symmetry properties.

When grouping by proximity is pitted against grouping by similarity, displays consist of at least two kinds of elements (called motifs), separated by one of the translation components, resulting in *dimotif lattices* (Figure 3; see Grünbaum & Shephard, 1987). In order to characterize the relation between two grouping principles, one must construct grouping *indifference curves* (Figure 4), similar to the indifference curves used in micro-economics (Krantz, Luce, Suppes, & Tversky, 1971): Imagine a consumer who would be equally satisfied with a market basket consisting of 1 kg of meat and 4 kg of potatoes and another consisting of 2 kg of meat and 1 kg of potatoes. In such a case, the (meat, potato) pairs (1, 4) and (2, 1) both lie on an indifference curve.

With these tools in hand, two important questions regarding perceptual grouping can be formulated. First, when several orientations can be perceived based on grouping by proximity in a particular dot lattice, what determines the preferred grouping? Is the outcome determined by the relative distance alone, or also by the angle between the competing organizations (an aspect that affects the global symmetry of the lattice and the way the overall configuration looks)? Second, when grouping by proximity and grouping by similarity are concurrently applied to the same pattern, what rule governs their joint application? Are these two principles combined additively or not? We briefly review the most important attempts to address these two questions. The initial studies always pitted grouping by proximity against grouping by similarity.

3.3.2 Initial attempts to quantify grouping by proximity by pitting it against similarity—

The first to systematically study grouping by proximity in interaction with similarity was Rush (1937). In her experiment she showed observers sequences of dot lattices in which the distance between dots in one orientation was held constant, and the distance between dots in another orientation was reduced from trial to trial. She assumed—incorrectly as argued below—that one could measure the strength of the two principles by finding their point of equilibrium, and concluded that “...it may be said that Similarity equals about 1.5 cm of Proximity” (p. 90).

Roughly two decades later, Hochberg and Silverstein (1956, unaware of Rush’s work, as a footnote in Hochberg & Hardy, 1960, attests) also set out to solve the problem of measuring the strength of grouping by similarity by pitting it against grouping by proximity. In manipulating luminance differences or distances between dots, they produced grouping indifference curves (Figure 4). Reanalysis showed that an additive combination of proximity and similarity described their results best. Unfortunately, the logic employed by Hochberg and his colleagues suffered from the same flaw as Rush. Their method produced only one grouping indifference curve—the one for which both groupings are in equilibrium. Their method cannot produce grouping indifference curves for which one principle is $2\times$ or $3\times$ as strong as the other, which are needed to measure the relative strengths of the two principles.

Quinlan and Wilton (1998) studied the relations between grouping by proximity and two forms of grouping by similarity (by color and by shape). Their stimuli consisted of strips of seven elements, with the center element as the target. They manipulated proximity by slightly shifting the left or right set of three elements, and they also manipulated color and

shape similarity. Observers were asked to rate the degree to which the target grouped with the elements on the left or on the right. Although the conception of the experiment is elegant, its reach was curtailed because each grouping principle was either present or absent. Had Quinlan and Wilton used a design in which each type of grouping was a multilevel factor, they could have addressed the additivity question (i.e., the second question introduced above), but they did not.

Oyama, Simizu, and Tozawa (1999) presented rectangular dimotif lattices for 3 s and asked observers to indicate continuously with a joystick whether they saw horizontal or vertical grouping. The horizontal separation was increased by 15° after a “horizontal” response, and decreased by that amount after a “vertical” response. Using a double-staircase method, the ratio of vertical to horizontal distances was determined that matched a particular dissimilarity. This method produced more complete equilibrium grouping indifference curves than were obtained by Hochberg and his colleagues.

3.3.3 The pure-distance law and the additivity of grouping principles—Oyama (1961) was the first to show that one can measure the strength of grouping by proximity without pitting it against another grouping principle (e.g., grouping by similarity). Using rectangular dot lattices at a fixed orientation, he recorded the amount of time subjects reported seeing the competing horizontal and vertical groupings. The ratio of the time they saw the horizontal and vertical organizations was found to be a power function of the ratio of the horizontal and vertical distances $t_h/t_v = (d_h/d_v)^{-\alpha}$, with $\alpha \approx 2.89$.

Using dot lattices at near-equilibrium, Kubovy and Wagemans (1995) and Kubovy, Holcombe, and Wagemans (1998) demonstrated that grouping by proximity can be understood as the outcome of a probabilistic competition among potential perceptual organizations. The basic idea is simple. If the distances between dots in two orientations of the lattice are equal, the chances of seeing one orientation or the other are equal too. If one distance becomes larger than the other, the relative chance of seeing that orientation decreases. If the ratio of the longer to the shorter vector is larger than about 1.5, grouping along that orientation is almost never seen. Kubovy and colleagues presented different kinds of dot lattices for 300 ms each and asked observers to indicate the perceived orientation. They could then use the frequencies of the perceived orientations over a large number of trials as estimates of the probabilities, and plot the relative frequencies as a function of relative distance. Their results (shown schematically in Figure 5) were remarkable. All the values of the log-odds fell on the same line, called the *attraction function*. Its slope is a person-dependent measure of sensitivity to proximity. Although the (relative) strength of grouping decays as an exponential function of (relative) distance, the attraction function in log-space is linear. The fact that all data points—obtained with all pairs of distances and all relative orientations (i.e., all points in the 2-D lattice space of Figure 2B)—could be fitted well by a single straight line indicates that grouping by proximity depends only on the relative distance between dots in competing organizations, not on the overall configuration in which the competition occurs (i.e., the lattice type, each with its own symmetry properties). Hence, this result, which was called *the Pure Distance law* provided a satisfactory answer to the first question raised above.

Once it has been established how grouping varies as a function of relative distance, the effect of conjoined grouping principles can be determined by measuring a family of grouping indifference curves. Kubovy and van den Berg (2008) presented participants with rectangular lattices of dots of different contrasts. Dots with the same contrast were either arranged along the shorter axis of each rectangle of dots within the lattice (similarity and proximity in concert) or arranged along the longer axis (similarity and proximity working against each other). Dot lattices varied across two dimensions: the ratio between the short

and long axis of each rectangle of dots within the lattice and the contrast difference between the different arrays of dots. As in the previous studies, each lattice was presented for 300 ms, and participants were asked to indicate which of the four orientations best matched the perceived arrangement of the dots in the lattice. By plotting the log likelihood of reporting the direction of the long axis versus the short axis as a function of the ratio of the length of the long and short axis for different values of the contrast difference between dots (shown schematically in Figure 6A), a family of grouping indifference curves was then obtained (depicted in Figure 6B). Because these indifference curves are parallel in log-odds space, the conjoined effects of proximity and similarity are called additive. Using lattices in which dots were replaced by Gabor elements, Claessens and Wagemans (2005) came to similar conclusions regarding proximity and collinearity. These results, therefore, provide a clear answer to the second question raised above.

3.4 Grouping Principles in Discrete Dynamic Patterns

Apparent motion is perceived when an object is presented in two or more successive frames at different spatial locations with proper durations and intervals. As discussed before, Wertheimer (1912) showed that under certain conditions it is possible to perceive pure motion, where motion is perceived without perceiving the moving object itself. The optimal timing and spacing between successively presented object presentations was investigated in more detail by Korte (1915), who found a direct relationship between the optimal temporal and the optimal spatial interval for perceived apparent motion. Later studies, however, have shown that the relationship between the optimal temporal and spatial interval depends on the stimuli used. For example, using horizontal arrays of dots that were displaced on successive frames, Burt and Sperling (1981) found that the spacing of the dots in the array strongly influenced the apparent motion percept in addition to the effects of the temporal interval between frames and the spatial displacement of the entire array.

The influence of spatial and temporal factors in apparent motion was further investigated using the Ternus display (Kramer & Yantis, 1997; Pantle & Picciano, 1976; Ternus, 1926; Wallace & Scott-Samuel, 2007), in which an array of three dots is presented across two frames at different spatial locations. When the two frames are presented in rapid succession (i.e., with a short inter-stimulus interval), it appears that the outmost dot is displaced while the center two dots appear to be stationary: *element motion* occurs. When the temporal interval between the successive frames is longer, the entire array of dots appears to jump: *group motion* is perceived. The two different types of perceived apparent motion represent two different solutions to the correspondence problem (Ullman, 1979), referring to the task of matching the objects in the first frame to the (possibly displaced) objects in the second frame. Whether element or group motion was perceived was found to depend on the properties of the individual stimuli in both frames such as their features (Dawson, Nevin-Meadows, & Wright, 1994), their size or the sharpness of their edges (Casco, 1990), as well as on the presence of contextual elements affecting how they are grouped (Kramer & Yantis, 1997) or how they are perceived in 3-D space (He & Ooi, 1999).

The interaction between spatial and temporal aspects was further investigated by Gepshtein and Kubovy (2000), who were able to determine the relationship between spatial grouping (determining which elements in each frame belong together) and temporal grouping (determining which elements across frames belong together), by using successive presentations of dot lattices—*motion lattices*—which allowed them to independently manipulate the strength of spatial and temporal groupings. A motion lattice (Figure 7) is composed of two identical dot lattices, D_1 and D_2 , displayed in alternation. Two ratios determine the perceived motion: (1) the motion ratio $r_m = m_2/m_1$, where m_1 and m_2 are the shortest and the next shortest spatial distances across which the apparent motion could occur between the frames; (2) the baseline ratio $r_b = b/m_1$, where b is the shortest spatial distance

between the dots within D_1 and D_2 to which the apparent motion could apply. The orientation of a virtual line drawn through these dots is called the baseline orientation.

As in the classic Ternus display, two classes of motion can be perceived. First, element motion is now apparent motion from each dot in D_1 to a corresponding dot in D_2 (and vice versa as the dot lattices alternate). The log-odds of seeing m_2 rather than m_1 as a function of the ratio of the distances is called an *affinity function*, by analogy with the concept of an attraction function for static dot lattices (Figure 8A). Second, group motion is now apparent motion orthogonal to the baseline orientation (Figure 7). Sequential models predict that if the spatial configuration of a stimulus remains constant, the likelihood of seeing group motion—an indicator of spatial grouping—cannot be affected by manipulations of the temporal configuration of the stimulus. However, the pattern of interaction in Figure 8B between r_m , the temporal configuration of the stimulus, and r_b , the relative density of the dots along the baseline, clearly refutes the sequential model.

How spatial and temporal distances interact to determine the strength of apparent motion has been controversial. Some studies report *space-time coupling*: If the spatial or temporal distance between successive stimuli is increased, the other distance between them must also be increased to maintain a constant strength of apparent motion (i.e., Korte's third law of motion). Other studies report *space-time trade-off*: If one of the distances is increased, the other must be decreased to maintain a constant strength of apparent motion. To establish what determines whether coupling or trade-off occurs, Gepshtein and Kubovy (2007) generalized the motion lattice of Figure 7, as illustrated in Figure 9, showing a temporal component of m_3 , T_3 , of twice the magnitude of the temporal component of m_1 , T_1 . By manipulating the spatial components of these motions, S_3 and S_1 from $S_3 \gg S_1$ to $S_3 \ll S_1$, an equilibrium point between the extremes was found at $r_{31} = S_3/S_1$, for which the probability of seeing the two motions was the same. If $r_{31} > 1$ then space-time coupling holds; if $r_{31} < 1$ then space-time trade-off holds. This suggests that previous findings on apparent motion were special cases and that the allegedly inconsistent results can be embraced by a simple law in which a smooth transition from trade-off to coupling occurs as a function of speed: Trade-off holds at low speeds of motion (below $\approx 12^\circ/\text{s}$), whereas coupling (Korte's law) holds at high speeds. The deeper theoretical implications of these results for the visual system's economy principles are discussed in the second review paper (Wagemans et al., 2012; Section 4).

The above research on perceptual grouping in static and dynamic discrete patterns spans a complete century, from Schumann (1900), Wertheimer (1912, 1923) and Korte (1915) up until today. It was mainly using well-controlled, parametrically varied stimuli in order to isolate one factor or another, and trying to quantify its strength. In addition, it sparked a renewed interest in understanding the level at which perceptual grouping operates, which was addressed in studies that use somewhat richer stimuli with additional variations, more typical for naturally occurring stimulation.

3.5 At What Level Does Grouping Happen?

As described above, Wertheimer (1923) demonstrated powerful grouping effects due to a large number of stimulus variables (e.g., proximity, similarity, good continuation) using flat 2-D displays on the printed page (see Figure 1). Subsequent researchers have investigated where in the visual system these effects occur (i.e., before or after the construction of a 3-D representation of the image), by using various kinds of 3-D displays with depth cues, shadows, transparency, and other higher-level factors.

Rock and Brosgole (1964) conducted a classic experiment on this topic to examine whether grouping by proximity operated on retinal 2-D distances or perceived 3-D distances.

Observers in a dark room saw a 2-D array of luminous beads either in the frontal plane (perpendicular to the line of sight) or slanted in depth so that the horizontal dimension of the array was foreshortened. The beads were actually closer together vertically than horizontally, so that when they were viewed in the frontal plane, observers always reported seeing them grouped into vertical columns rather than horizontal rows. The critical question was whether or not the beads would be grouped in the same way when the same lattice was viewed slanted in depth such that the beads were retinally closer together in the horizontal direction. When this array was viewed monocularly, so that the beads appeared to be in a frontal plane perpendicular to the line of sight (even though they were actually slanted in depth), observers perceived the grouping to change to a set of rows rather than columns, as one would expect based on retinal distances. However, when viewed binocularly, so that stereoscopic depth information enabled observers to see the beads slanted in depth, they reported grouping them into vertical columns, as predicted by postconstancy grouping based on a 3-D representation of perceived distances in the phenomenal environment (because the beads appeared to be closer in the vertical direction, as was actually the case in the physical world). Rock and Brosdale's results therefore support the hypothesis that the final, conscious result of grouping occurs after binocular depth perception. Several phenomenological demonstrations supporting the same conclusion are provided by Palmer (2002b; Palmer, Brooks & Nelson, 2003).

Rock, Nijhawan, Palmer and Tudor (1992) later investigated whether grouping based on lightness similarity happened before or after lightness constancy. Using displays that employed cast shadows and translucent overlays, they also found evidence that the final conscious result of grouping depended on a postconstancy representation that reflected the perceived reflectance of surfaces rather than the luminance of retinal regions. Analogous evidence that the final conscious organization resulted from a grouping process that operates on relatively late, postconstancy representations was reported by Palmer, Neff, and Beck (1996) for amodal completion and by Palmer and Nelson (2000) for illusory contours. Further results of Schulz and Sanocki (2003) support the view that prior to achieving the conscious result of perceptual grouping based on a 3-D postconstancy representation, some nonconscious grouping processes operate on a 2-D preconstancy representation. They used the same lightness displays as did Rock et al. (1992), but included a brief, masked presentation condition in which they found that observers reported seeing an organization that is based on the retinal luminance of 2-D regions (see also van den Berg et al., 2011). Further evidence that grouping operations occur before constancy has been achieved is based on grouping effects that actually influence the achievement of constancy (see Palmer, 2003).

Perhaps the most parsimonious view consistent with the known facts is that grouping principles operate at multiple levels. It seems most likely that provisional grouping takes place at each stage of processing, possibly with feedback from higher levels to lower ones, until a final, conscious experience arises of a grouping that is consistent with the perceived structure of the 3-D environment. Whereas the above findings provide valuable information about the stages at which grouping operates, these studies have mainly employed relatively artificial stimuli. The next section is dedicated to the role of grouping in contour integration and completion, in ways that are closer to the processing of natural stimuli.

4 Contour Integration and Completion

4.1 Introduction

Studies in which grouping factors are isolated to quantify their strength are useful but understanding their role in everyday perception requires a different approach. An important task of natural vision is to identify and group together the portions of the 2-D retinal image

that project from an object. In the simple case in which the object boundary projects as a single closed curve, the problem reduces to a problem of contour grouping or contour integration. From the fifty-year history of computer vision research, however, we know that this is a computationally difficult problem for a number of reasons (e.g., Elder, Krupnik, & Johnston, 2003). First, occlusions occur generically in natural images, resulting in a projection of the rim of the object as a disconnected set of contour fragments (see Figure 10). Also, where the figure-ground contrast is low, portions of the contour may not be detected, resulting in further fragmentation. To complicate matters further, natural images are often highly cluttered, such that for any given contour fragment, multiple other fragments could be the correct continuation of the contour. Thus, to effectively exploit contours for object segmentation, the visual system must be able to cope with uncertainty, using a relaxed form of perceptual contour closure that can work reliably even for fragmented contours (e.g., Elder & Zucker, 1993). Discovering the nature of these mechanisms is a central topic in modern perceptual organization research, involving a range of different methodologies, including psychophysics, neurophysiology, neuroimaging, ecological statistics, and computational modeling.

Addressing the role of grouping under natural conditions requires a definition of the grouping primitives: What exactly are the elements being grouped? In the research with dot lattices reviewed above, the primitives are zero-dimensional points or dots, and proximity and similarity were the only cues involved. The study of contour grouping over oriented primitives such as bars or edges expands the grouping cues to also include good continuation. This maps directly onto related computer vision research, where the contour grouping problem is typically defined over local oriented edge elements detected using oriented linear filters (e.g., Canny, 1983) that can be loosely identified with receptive fields of simple cells in primate V1 (Hubel & Wiesel, 1968). Even higher-level primitives can be used, for instance, connected contour fragments of arbitrary shape, which may have been separated by occlusions. This higher level of representation may play a role in perceptual completion (see also below), and could be mapped to extrastriate visual areas such as V4, where neurons are known to be selective for higher-order properties of shape (Pasupathy & Connor, 1999). In psychophysics, this has motivated the use of fragmented object pictures, which can be used to study the dynamic interplay between perceptual grouping and object identification (e.g., Panis & Wagemans, 2009).

In the following subsections, we first review the research on the role of grouping principles for contour integration—research that generalizes properties discovered for individual cues in isolation to more natural conditions where multiple cues are present and the visual stimulus may be quite complex. In addition to psychophysical results, we review the ecological foundations of the problem, and discuss computational principles and possible neural mechanisms. We then turn to the specific problem of contour completion in cases of occlusion, before discussing general issues pertaining to both perceptual grouping and contour integration.

4.2 Grouping Principles for Contour Integration

4.2.1 Proximity—The principle of proximity states that the strength of grouping between two elements increases as these elements are brought nearer to each other, but how exactly does grouping strength vary as a function of their separation? As reviewed above, Oyama (1961) found that this relationship could be accurately described as a power law, whereas Kubovy and Wagemans (1995) employed an exponential model, consistent with random-walk models of contour formation (Mumford, 1992; Williams & Jacobs, 1997). However, Kubovy et al. (1998) also noted that a power law model could fit their data equally well and found that the proximity cue was approximately scale-invariant: Scaling all

distances by the same factor did not affect results. Since the power law is the only perfectly scale-invariant distribution, this last result adds strength to the power-law model of proximity, which has been used in subsequent studies (e.g., Claessens & Wagemans, 2008).

Perceptual scale invariance is rational if in fact the proximity of elements along real contours in natural images is scale invariant—that is, if the ecological distribution follows a power law. In support of this idea, Sigman, Cecchi, Gilbert, and Magnasco (2001) reported that the spatial correlation in the response of collinearly-oriented filters to natural images does indeed follow a power law. Quantitatively, however, the correspondence between psychophysics and ecological statistics is poor here: While Oyama estimated the perceptual power law exponent to be $\alpha \approx 2.89$, Sigman et al. estimated an ecological exponent of only 0.6, reflective of a much weaker cue to grouping. However, Sigman et al. (2001) did not restrict their measurements to pairs of neighboring elements on the same contour of the image. In fact, the measurements were not constrained to be on the same contour, or even on a contour at all, leading to a mixture between strongly-related and only weakly-related image features. Elder and Goldberg (2002) estimated these distributions more directly, asking human observers to label the sequence of elements forming the contours of natural images, with the aid of an interactive image editing tool. This technique allowed the measurements to be restricted to successive elements along the same contour, and yielded a clear power law with exponent $\alpha = 2.92$, very close to the perceptual estimate of Oyama. Whether this exponent is independent of orientation in the image remains an interesting open question, but psychophysical data (Claessens & Wagemans, 2008) suggest that at least perceptually this may not be the case.

In sum, the convergence between psychophysics and ecological statistics is compelling: Ecologically, proximity follows a power law and exhibits scale invariance, and these properties are mirrored by the psychophysical results. Thus, we have a strong indication that the human perceptual system is optimally tuned for the ecological statistics of proximity cues in natural scenes.

4.2.2 Good continuation—A second important grouping principle for contour integration is good continuation, which refers to the tendency for elements to be grouped to form smooth contours (Wertheimer, 1923/1938). A very effective method for studying the principle of good continuation in cluttered images was developed by Field, Hayes, and Hess (1993). In this method, a contour formed from localized oriented elements is embedded in a random field of homogeneously distributed distractor elements, in order to eliminate the role of proximity (Figure 11). Aligning the contour elements tangentially to the contour makes the contour easily detected, whereas randomizing the orientation of the elements renders the contour invisible, clearly demonstrating the role of good continuation in isolation from proximity.

These findings led Field et al. (1993) to suggest the notion of an *association field* that determines the linking of oriented elements within a local visual neighborhood (Figure 12), a construct that is closely related to the machinery of cocircularity support neighborhoods, developed earlier for the purpose of contour refinement in computer vision (Parent & Zucker, 1989).

Ecological data on good continuation have also emerged over the last decade. Kruger (1998) and later Sigman et al. (2001) found evidence for collinearity, cocircularity and parallelism in the statistics of natural images. Geisler, Perry, Super, and Gallogly (2001) found similar results using both labeled and unlabeled natural image data, in fairly close correspondence with the tuning of human perception to the good continuation cue. Geisler et al. (2001) treated contours as unordered *sets* of oriented elements, measuring the statistics for pairs of

contour elements on a common object boundary, regardless of whether these element pairs were close together or far apart on the object contour. In contrast, Elder and Goldberg (2002) modeled contours as ordered *sequences* of oriented elements, restricting measurements to adjacent pairs of oriented elements along the contours. The likelihood ratios for two oriented elements to be neighboring elements on the same object boundary are much larger for the sequential statistics, reflecting a stronger statistical association between neighboring contour elements.

4.2.3 Similarity—The principle of similarity states that elements with similar properties (e.g., brightness, contrast, color, texture) are more likely to group than elements that differ on these dimensions, which has been demonstrated in a number of ways with dot patterns (see above). For oriented elements, studies have generally found a decline in contour integration performance for contrast reversals under certain conditions but not others, suggesting non-additive interactions with other grouping cues such as proximity and good continuation (e.g., Elder & Zucker, 1993; Field, Hayes, & Hess, 2000; Gilchrist, Humphreys, Riddoch, & Neumann, 1997; Rensink & Enns, 1995; Spehar, 2002). Elder and Goldberg (2002) explored the ecological statistics of similarity in edge grouping, coding similarity in terms of the difference in brightness and in contrast between the edges, and found that the brightness cue carries useful information for grouping but the contrast cue is relatively weak. Whereas Elder and Goldberg restricted their study to pairs of elements of the same contrast polarity, Geisler and Perry (2009) have more recently studied the ecological statistics of contrast polarity, demonstrating that it is also an informative cue for contour integration.

4.2.4 Closure—The role of closure in contour integration has been debated. Whereas Kovács and Julesz (1993) found superior detection performance for closed, roughly circular contours, compared to open curvilinear controls, these findings might also be based, in part, on good continuation alone. In fact, with stimuli more closely controlled for good continuation cues, Tversky, Geisler, and Perry (2004) found only a small advantage for closed contours and argued that this advantage could be due to probability summation rather than closure. While this negative result may seem inconsistent with the emphasis placed on closure in early Gestalt work, it is important to appreciate the exact nature of this early view, as expressed by Koffka (1935, p. 150): “Ordinary lines, whether straight or curved, appear as lines and not as areas. They have shape, but they lack the difference between an inside and an outside... If a line forms a closed, or almost closed, figure, we see no longer merely a line on a homogeneous background, but a surface figure bounded by the line.” The original Gestalt claim was thus not that closure is a grouping cue per se, but rather that it somehow profoundly determines the final percept of form. In the same spirit, Elder and Zucker (1993, 1994) argued that the most important role for closure was as a bridge from 1-D contour to 2-D shape, a suggestion that was supported by the finding that small changes in good continuation and closure can yield large changes in shape discriminability.

4.2.5 Symmetry and parallelism—The Gestaltists identified *symmetry* as a factor of “good shape” (Koffka, 1935), although it seems to be easily overruled by good continuation and convexity (Kanizsa, 1979). In the computer vision literature, symmetry has been used in numerous contour integration algorithms (e.g., Mohan & Nevatia, 1992; Stahl & Wang, 2008). *Parallelism* has been identified as a factor determining the perceptual simplicity of line configurations (Arnheim, 1967), and as a grouping cue in computer vision algorithms (e.g., Jacobs, 2003; Lowe, 1985). In the ecological statistics literature, parallelism has been studied extensively as a cue for the grouping of oriented edge elements into contours (Elder & Goldberg, 2002; Geisler et al., 2001; Kruger, 1998). Psychophysical evidence for the role of symmetry and parallelism in contour integration has been reported by Feldman (2007),

who showed that comparison of features lying on pairs of line segments is significantly faster if the segments are parallel or mirror-symmetric, suggesting a fast grouping of the segments based upon these cues. More recently, Machilsen, Pauwels, and Wagemans (2009) demonstrated enhanced detectability of bilaterally symmetric versus asymmetric closed forms, suggesting a role for more complex, global symmetry processing in contour integration. Physiologically, it is known that bilaterally symmetric patterns differentially activate human extrastriate visual areas V3, V4, V7 and LO, and homologous areas in macaque cortex (Sasaki, 2007).

4.2.6 Convexity—Convexity has long been known as a grouping cue (Rubin, 1927). In the computer vision literature, Jacobs (1996) also demonstrated its utility for grouping contour fragments that can then be used as features for object recognition. Liu, Jacobs, and Basri (1999) subsequently employed a novel psychophysical method to demonstrate the role of convexity in contour integration. Their method relies on the observation of Mitchison and Westheimer (1984) that judging the relative stereoscopic depth of two contour fragments becomes more difficult when the fragments are arranged to form a configuration with good continuation and closure. Using an elaboration of this method, Liu and colleagues showed that stereoscopic thresholds are substantially higher for occluded contour fragments that can be completed to form a convex shape, relative to fragments whose completion induces one or more concavities. This suggests that the visual system is using convexity as a grouping cue.

In sum, research on contour integration, using oriented elements or shape fragments, has demonstrated important roles for the principles of proximity, good continuation, similarity, closure, symmetry and parallelism. This research thus can be seen as an elaboration of early Gestalt work that established these principles phenomenologically. Importantly, however, work on contour integration has provided quantitative measures which have led directly to computational models and algorithms, and which have allowed direct comparison to the underlying ecological statistics of the problem and to neurophysiological results.

Whereas the preceding section discussed research on the integration of contour fragments, the next section turns to the specific problem of contour completion in cases of occlusion, which brings us another step closer to understanding the role of grouping in more complex tasks of natural vision.

4.3 Contour Completion

4.3.1 Modal and amodal completion—Whereas contour integration refers to the integration of discrete elements such as dots or oriented elements, contour completion refers to the integration of smooth extended contours, separated by gaps due to occlusion or camouflage. Stimuli used in the study of contour completion often include cues to relative depth or occlusion, such as T-junctions (see Figure 13). In Figure 13A, for example, the perceived unity of the black fragments and the perceived shape of the black surface behind the gray occluder appear to result from visual contour completion. This form of completion of the black shape behind the gray occluder is referred to as *amodal completion* (Michotte et al., 1964). Although one has a compelling sense of continuity of the boundaries of the black surface behind the occluder, one does not actually see a contour. By contrast, illusory contours constitute an example of *modal completion*. They generate a percept of a contrast border in image regions that are physically homogeneous, such as the illusory edges of the white wedge in Figure 13B.

4.3.2 Grouping and shape problem—For contour completion, the visual system must solve two problems (see Figure 14). First, it must determine whether or not the two contour

fragments are part of a single continuous contour: the *grouping problem*. Second, it must determine the shape of the missing portion of the contour: the *shape problem*.

As a solution to the grouping problem, Kellman and Shipley (1991) have proposed a relatability criterion. This criterion refers to the coding of contour geometry in terms of the positions and orientations of the edges of the contours. Two inducing contours are grouped if they satisfy two conditions. First, if extended linearly, these extensions must intersect and, second, the turning angle of the two edges should not exceed 90°. Evidence for the role of contour relatability has been found by means of observer ratings (Kellman & Shipley, 1991), shape (fat/thin) discrimination performance (Ringach & Shipley, 1996; Zhou, Tjan, Zhou, & Liu, 2008), depth discrimination performance (Liu et al., 1999; Yin, Kellman, & Shipley, 2000), and interpolation settings (Fulvio, Singh, & Maloney, 2008).

The shape problem involves determining the shape of the occluded portion of the contour. This is a highly under-constrained problem, since an infinite number of possible shapes could be hidden behind the foreground shape (see Figure 14C). Human observers typically perceive only a small subset of these. This suggests that the visual system imposes strong constraints when generating and selecting the set of possible shapes. Different methods have been used to examine how the visual system solves the shape problem. For example, several studies have asked participants to draw the perceived trajectory of visually completed contours (Takeichi, Nakazawa, Murakami, & Shimojo, 1995; Ullman, 1976). Other studies have asked observers to mark the farthest point along the interpolated contours (Fantoni, Bertamini, & Gerbino, 2005; Guttman & Kellman, 2004; Takeichi et al., 1995), to match the perceived interpolated shape from a parametric family (Singh, 2004), or to rate its overall degree of “roundedness” (Fantoni & Gerbino, 2003). More recently, Fulvio et al. (2008) used a method that obtains interpolation settings (position and orientation) at multiple locations along the length of a partly-occluded contour. An examination of the precision and consistency of these settings showed that observers can interpolate a single, stable (self-consistent), smooth contour when the inducing contours are relatable, but not when they require an inflecting interpolating curve.

An empirically successful model which addresses the grouping problem and the shape problem simultaneously has been proposed by van Lier (van Lier, 1999; van Lier, van der Helm, & Leeuwenberg, 1994, 1995). In line with the descriptive simplicity principle (see Wagemans et al., 2012; Sections 5 and 6), this model postulates that an interpretation is judged not only by the complexities of the potentially occluding and occluded shape(s) but also by the complexity of the relative positions of these shapes. The combination with the lowest overall complexity is then taken to predict whether occlusion occurs, and if so, how the occluded shape looks like.

4.3.3 Contour interpolation and extrapolation—Good continuation is an important criterion in shape interpolation, but it is not perfectly clear what is meant by this term. Dating back to Wertheimer (1923, especially his Figures 16–19), two aspects have been distinguished: (1) what geometric properties of the visible contours are used by human vision (e.g., orientation, curvature, rate-of-change of curvature, higher-order derivatives) and (2) how are they combined to define the shape of the contour?

These questions were addressed in studies of *contour extrapolation* (e.g., Singh & Fulvio, 2005, 2007), in which observers adjusted the position and orientation of a line probe on the other side of an occluder (Figure 15) at multiple distances from the point of occlusion in order to optimize the percept of smooth continuation. The findings suggested that observers use contour orientation and curvature, but *not* rate-of-change of curvature. The shape of visually extrapolated contours was best characterized by a decaying curvature. The results

were modeled by a Bayesian interaction between a tendency to extend contour curvature and a prior tendency for straight contours.

Models of *contour interpolation* typically use just the position and orientation of the visible segments at their respective points of occlusion, but not their curvature (e.g., Fantoni & Gerbino, 2003; Kellman & Shipley, 1991; Ullman, 1976). The above results suggest, however, that interpolation models should use curvature as well, especially if they aim to model human contour interpolation. How much of the visible contours the visual system uses is still an open question. More flexible behavior can be obtained (in terms capturing various dependencies exhibited by human vision) using probabilistic models of contour interpolation rather than those based on fixed classes of shapes, or that minimize fixed geometric criteria (see Singh & Fulvio, 2007).

4.3.4 Surface geometry and layout—The dominant approach in contour completion has been to treat the problem in isolation (e.g., Kellman & Shipley, 1991), but there is growing evidence that visual contour completion is very much informed and guided by surface geometry and surface layout.

One aspect of *surface geometry* is curvature polarity, where a curved contour segment (Figure 16A) can correspond to a locally convex (Figure 16B) or a concave portion of a surface (Figure 16C). In addition, curvature on an object boundary can arise either because the axis of the object itself is curved (like a bending “ribbon”; Figure 16D), or because of the changing width of the shape around a straight axis (a convex “bulge” or a concave “neck”; Figure 16B–C). Experimental studies have shown that the convexity of the surface influences grouping in both contour completion (Liu et al., 1999; Tse, 1999) and interpolated shape (Fantoni et al., 2005; Fulvio & Singh, 2006). Moreover, the influence of surface geometry goes beyond local convexity to include skeletal shape description (Fulvio & Singh, 2006).

In addition to surface geometry, *surface layout* also exerts an influence on contour completion, for example, by differences in perceived depth. In amodal completion, the completed contour tends to belong to the occluded surface. In modal completion, in contrast, it belongs to a nearer, occluding surface. Such effects of surface layout are not consistent with a single mechanism underlying both modal and amodal completion (Anderson, Singh, & Fleming, 2002; Zhou et al., 2008). Instead of assuming two separate mechanisms for modal and amodal completion, however, it could also be assumed that the contour interpolation mechanism is flexible, in that it can take into account various image conditions (e.g., relative depth, photometric conditions, as well as surface geometry; Singh, 2004; Zhou et al., 2008). The exact mechanism underlying modal and amodal completion has been the topic of much recent debate (e.g., Anderson, 2007; Kellman, Garrigan, Shipley, & Keane, 2007).

In sum, when we explicitly study contour completion under occlusion, some additional issues arise, and research indicates that the mechanisms take more complex geometric aspects into account, that they are more sensitive to the configuration in which the visible contour fragments occur, and that they are generally more flexible, than in cases of contour completion without occlusion.

4.4 Some General Issues Regarding Perceptual Grouping and Contour Integration

While our review has thus far considered each individual Gestalt principle in turn and how they apply to contour integration and completion, there are key overarching questions that cut across all of these cases of perceptual grouping: (1) To what extent are the Gestalt laws

innate or learned?; (2) how are they combined?; and (3) how can they be jointly represented in accurate computational models and useful algorithms?

4.4.1 Development—In general, Gestalt psychology has tended to emphasize the degree to which the Gestalt laws are innate or intrinsic to the brain rather than learned from past experience. Research suggests that infants are capable of grouping visual elements into unitary structures in accord with a variety of both classic and modern organizational principles (for a review, see Bhatt & Quinn, 2011). Infants as young as 3 to 4 months of age show grouping by good continuation, proximity, connectedness, and common region, and grouping by lightness similarity was observed even in newborns. However, only 6- to 7-month-olds appear to utilize form similarity to organize visual patterns (Quinn & Bhatt, 2006), and 3-month-olds appear to be completely insensitive to closure (Gerhardstein, Kovács, Ditre, & Feher, 2004), suggesting that not all grouping cues are readily available to young infants.

Furthermore, recent psychophysical studies of older children suggest that there is a protracted developmental trajectory for some perceptual organization abilities, even those that appear to emerge during infancy (e.g., Hadad & Kimchi, 2006; Hadad, Maurer, & Lewis, 2010a, 2010b; Kimchi, Hadad, Behrmann, & Palmer, 2005; Kovács, 2000). For example, the Gestalt principles underlying contour grouping continue to develop in children through their late teens. These findings suggest that visual experience plays a role in at least some aspects of perceptual organization. Further support for this notion comes from studies demonstrating influence of perceptual learning on organizational abilities in infancy (e.g., Quinn & Bhatt, 2005), and the influence of associative learning (Vickery & Jiang, 2009), and past experience (Kimchi & Hadad, 2002) on perceptual grouping in adult observers.

4.4.2 Cue combination—One of the central questions in grouping concerns the way in which the brain combines multiple cues to yield a unitary organization. Historically, this problem has often been posed in terms of competitive interactions formulated either in descriptive terms (usually seeking compliance with the simplicity principle) or in probabilistic terms (usually Bayesian formulations which may or may not seek compliance with the Helmholtzian likelihood principle). (For more details on this, see Wagemans et al. (2012), Sections 5 and 6.) However, in natural scenes, disparate weak cues can often combine synergistically to yield strong evidence for a particular grouping.

To explore this issue, Geisler et al. (2001) used a nonparametric statistical approach, jointly modeling the ecological statistics of proximity and good continuation cues as a 3-D histogram, to show that human observers combine these two classic Gestalt principles in a roughly optimal way. Elder and Goldberg (2002) demonstrated that the ecological statistics of proximity, good continuation, and similarity are roughly uncorrelated, so that to a first approximation the Gestalt laws can be factored: The likelihood of a particular grouping can be computed as the product of the likelihoods of each grouping factor in turn (or, equivalently, the log likelihood of the grouping factor is the sum of the log likelihoods of the grouping factors; for further explanation, see Claessens & Wagemans, 2008).

Elder and Goldberg's (2002) approach also allowed quantification of the statistical power of each Gestalt cue, as the reduction in the entropy of the grouping decision given each cue individually. They found that proximity was by far the most powerful, reducing the entropy by roughly 75%, whereas good continuation and similarity reduced entropy by roughly 10% each. The most accurate grouping decisions could be made by combining all of the cues optimally according to the probabilistic model, trained on the ecological statistics of natural images. Such a statistically optimal combination of grouping cues has also received some psychophysical support (Claessens & Wagemans, 2008).

4.4.3 Computational models—Numerous computer vision algorithms for grouping make use of the Gestalt factors described above (e.g., Elder et al., 2003; Estrada & Elder, 2006; Jacobs, 1996; Lowe, 1985; Sha’ashua & Ullman, 1988; Stahl & Wang, 2008). Most of these use the local Gestalt principles of proximity, good continuation, and similarity to group long chains under an explicit or implicit Markov assumption, while additional global factors of convexity (Jacobs, 1996), closure (Elder et al., 2003), and symmetry (Stahl & Wang, 2008) have been employed to further condition the search. In addition, coarse-to-fine feedback techniques have been used as an alternative way to incorporate more global constraints into the grouping process (Estrada & Elder, 2006). Grounding probabilistic algorithms in the ecological statistics of contour grouping avoids the ad hoc selection of algorithm parameters and optimizes performance on natural scenes (Elder et al., 2003; Estrada & Elder, 2006).

Although the Gestalt principles of grouping were largely based on the analysis of figures in the 2-D image plane, more recent work derives these principles from the geometric laws of 3-D projection, within the theoretical framework of minimal viewpoint invariants (Jacobs, 2003; Lowe, 1985). Briefly, the theory is based upon the assumption that the observer takes a general viewpoint position with respect to scenes. This assumption implies that certain so-called non-accidental properties in the 2-D proximal stimulus are most likely properties of the 3-D distal stimulus as well (Binford, 1981; Lowe, 1985). Examples of such non-accidental properties are the geometric Gestalt principles relevant to contour grouping: proximity, good continuation, closure, convexity, parallelism, and symmetry (e.g., Wagemans, 1992, 1993). This notion has played a central role as a bridge between the Gestalt grouping principles and object representation and recognition (Biederman, 1987).

4.4.4 Conclusion—Research on principles of perceptual grouping has come a long way since Wertheimer’s (1923) phenomenological demonstrations. Using well-controlled discrete patterns, grouping by proximity could be quantified as a real law. In the last two decades or so, new grouping principles have been identified, and their functional role in the processing of more complex images occurring in real-world situations became a central focus. In much of this research, quantitative psychophysical work went hand-in-hand with the analysis of ecological statistics, the development of computational models and neurophysiological research. In Section 5, we see a similar evolution in research on the closely related problem of figure-ground organization. Section 6 is devoted to an integrated review of the research on neural mechanisms of contour grouping and figure-ground organization.

5 Figure-Ground Organization

5.1 Introduction

Two adjoining regions of the visual field with a shared border can lead to a mosaic percept, but more often it leads to figure-ground organization. In this percept, the shared border is perceived as the occluding edge of one of the regions. The occluding region is perceived as the figure, and it appears shaped by the border. The adjoining region appears to simply continue behind it as its background, with no shape being imparted by the border. The figure is said “to own the borderline.” When figure-ground reversals occur, the border-ownership switches as well (e.g., the famous vase-faces figure by Rubin, 1915). Why switching occurs and how its temporal characteristics inform us about the dynamics of the brain is discussed extensively in the second paper (Wagemans et al., 2012; Section 3). Here, we focus on the factors or principles that determine what is perceived as figure.

The question of whether or not experience plays a role has been particularly controversial. On the one hand, structuralists like Wundt argued that past experience was the sole

determinant of which region of the visual field was perceived as figure. On the other hand, Gestalt psychologists like Wertheimer, Koffka, and Köhler, considered the structuralist position untenable because it would draw too strongly on memory to be useful for real-time perception. They also pointed out that novel objects can be perceived easily, which is inconsistent with a theory in which only past experience causes perceptual organization. The Gestalt view was that figure-ground segregation occurs by innate, intrinsic segregation laws that embody aspects of the world in which human perception evolved. The resultant shaped entities (figures) are then matched with traces in memory, a process that requires a less extensive memory search. In this view, familiar and novel objects are distinguished only after they attain figural status.

Wertheimer argued that past experience was unlikely to influence *initial* organization. He put forward two requirements to be met for past experience to convincingly play a role in figure-ground perception (Wertheimer, 1923/1938, p. 86): “[The] duty [of the doctrine of past experience] should be to demonstrate ... (1) that the dominant apprehension [perception] was due to earlier experience (and to nothing else); (2) that nondominant apprehensions in each instance had not been previously experienced.” The Gestalt psychologists demonstrated that a variety of image properties were sufficient for the perception of figures or configurations, without the need of familiarity. These image properties—convexity, symmetry, small area, and “surroundedness” (or enclosure)—became known as the classic configural principles of figure-ground organization (e.g., Harrower, 1936; Rubin, 1915).

5.2 Classic Image-Based Configural Principles of Figure-Ground Organization

Traditionally, investigators have tested the effectiveness of configural factors in controlled experiments, using displays composed of multiple alternating black and white regions where either the black or the white region could have the configural factor (e.g., convexity in Figure 17). Participants were asked to indicate which region, the black or the white one, was perceived as the figure. These experiments support less formal early demonstrations of the importance of configural factors for initial figure-ground segregation, without the need to rely on past experience (familiarity). The most powerful demonstrations were obtained with convexity and symmetry, although these principles were often confounded in earlier studies (e.g., Bahnsen, 1928). Confounding often occurred in tests of the other two configural principles—small region and surroundedness—as well, which created some confusion on the individual effectiveness of each of these principles. Note that long exposure times were used in these early demonstrations, so it is not clear that reports reflected initial organization. Later, more controlled investigations of convexity and symmetry revealed that their strength was overestimated by traditional demonstrations.

Convexity—Also using long exposures, Kanizsa and Gerbino (1976), tested the effectiveness of convexity as a configural figure-ground principle, using displays similar to Figure 17, and found that regions with convex parts were seen as figure on approximately 90% of trials. These results—in tandem with the previous Gestalt demonstrations—were taken as evidence that part convexity might determine to a large extent the position of figures in the input array, even before memories of previously seen objects are accessed (see Singh & Hoffman, 2001). To examine whether the effects of part convexity are general in that they extend to stimuli with fewer regions than used in the original displays, Peterson and Salvagio (2008) presented displays that could contain either two, four, six, or eight regions. In addition, the exposure duration was reduced to 100 ms. Interestingly, the likelihood of perceiving convex regions as figures increased from only 57% of trials with two-region displays, up to 89% for eight-region displays (as in Kanizsa & Gerbino, 1976). This effect of the number of alternating regions—the convexity context effect—was

obtained only when the concave regions were homogeneous in color, whereas the color of the convex regions was irrelevant. Thus, homogeneity of alternating regions functions as a background cue, but other experiments indicated that it does so only when a configural cue favors perceiving the regions in-between as figures. These experimental observations are consistent with a Bayesian ideal observer model that estimates the probability that a picture depicts a 3-D scene or a 2-D pattern (Goldreich & Peterson, 2012). Specifically, to be able to fit Peterson and Salvagio's (2008) data, the Bayesian observer had weak biases for convex over concave objects and for single-color over multi-color occluded objects.

Symmetry—Kanizsa and Gerbino (1976) measured the effectiveness of global symmetry as a configural figure-ground principle by placing it in competition with convexity. They found that convex-region-as-figure reports were unaffected by competition from symmetry, suggesting that symmetry is only a weak factor (see also Pomerantz & Kubovy, 1986). However, with longer exposures of eight-region displays consisting of alternating symmetric and asymmetric regions (with no competition from convexity), Driver, Baylis, and Rafal (1992) found that symmetric regions were perceived as figures on a large proportion of the trials. When Peterson and Gibson (1994a) examined the effectiveness of symmetry as a configural figure-ground principle in two-region displays with brief exposure durations, their observers reported perceiving the symmetric region as figure on approximately three-quarters of the trials. Effects of global symmetry have also been found for shapes defined by a contour of Gabor elements presented in a noisy background (Machilsen et al., 2009), suggesting that global symmetry is effective in segregating figures from backgrounds. Together, these findings are consistent in showing that—at least near fixation—global symmetry is a configural figure-ground principle.

5.3 New Image-Based Principles of Figure-Ground Organization

All of the above classic figure-ground principles concern the organization of displays consisting of static, homogeneously colored regions. Quite a few additional principles of figure-ground organization have since been discovered, some of which also apply to static, homogeneously colored regions (e.g., lower region, and top-bottom polarity). Additional figure-ground principles come into play in displays containing spatial heterogeneities such as texture (extremal edges and edge-region grouping) and when they contain motion (advancing region, articulating concavities, and convex motion). Pinna (2010) recently even extended the range of application of these principles to shape and meaning. As with the new grouping principles, researchers have often provided an ecological foundation of these new figure-ground principles too.

Lower region—Vecera, Vogel, and Woodman (2002) showed that when a rectangular display is divided in half by an articulated horizontal border, the region below the border is more likely to be perceived as the closer, figural region than is that above the border. They found this lower-region bias most strongly in “city-scape” images in which the border consisted of horizontal and vertical segments, but also in analogous “stalactite/stalagmite” images consisting of curved segments. Vecera (2004) performed additional experiments in which such displays were viewed by observers whose heads were tilted (or even inverted) to determine whether this figure-ground bias was driven by a viewer-centered, retinal reference frame or an environmental reference frame. The results showed—somewhat surprisingly—that retinal directions were clearly dominant. Although this result is at odds with the presumed ecological justification of gravity as the rationale for perceiving lower regions as closer (see Vecera & Palmer, 2006), it is consistent with the need to compute information about figure-ground status early in visual processing, before orientation constancy has occurred. Indeed, because head orientation is approximately vertical most of the time, the difference between retinal and environmental reference frames is negligible. The ecological

validity of lower region was assessed statistically by analyzing a corpus of photographic images that were hand-segmented by human observers (Martin, Fowlkes, Tal, & Malik, 2001). The results showed that lower region was a valid cue to closer surfaces at local edges whose orientation was roughly horizontal. Vecera and Palmer (2006) argued that the effect of seeing lower regions as closer resulted from biases in the ecological statistics of depth edges at various orientations when objects rest on a ground plane beneath them, as do most terrestrial objects in the earth's gravitational field.

Top-bottom polarity—Hulleman and Humphreys (2004) showed that regions that are wider at the bottom and narrower at the top are more likely to be perceived as figures than regions that are wider at the top and narrower at the bottom. The regions in their displays looked a bit like odd evergreen trees or chess pieces, but they argued that their effects were not due to the effects of familiar shape (e.g., Peterson & Gibson, 1994a) because there are other familiar objects (e.g., tornados) that are similar in shape to the regions with narrow bases and wide tops. They also claim that top-bottom polarity effects cannot be explained by the effects of lower region (Vecera et al., 2002). Nevertheless, the ecological factor that links all three of these figure-ground factors (canonical orientation of familiar shapes, lower region, and top-bottom polarity) is gravity. Indeed, top-bottom-polarity can easily be interpreted as due to the perceptual consequences of gravitational stability.

Extremal edges and gradient cuts—An extremal edge (EE) in an image is a projection of a viewpoint-specific horizon of self-occlusion on a smooth convex surface, such as the straight side of a cylinder. Computational analyses of the visual properties of surfaces near EEs show characteristic gradients of shading and/or texture in which the equiluminance and/or equidensity contours are approximately parallel to the edge of the surface (Huggins & Zucker, 2001). EEs are relevant to figure-ground determination because the side with an EE gradient is almost invariably perceived as being closer to the observer than the opposite side of the edge (Palmer & Ghose, 2008), even when EE is placed in conflict with other factors (Ghose & Palmer, 2010).

Edge-region grouping—Palmer and Brooks (2008) built a bridge between classic grouping effects and figure-ground organization by reasoning that if figure-ground organization is indeed determined by an edge “belonging to” (i.e., grouping with) the region on one side more strongly than that on the other, then any grouping factor that could relate an edge to a region should also operate as a figure-ground factor. They tested this hypothesis for six different grouping factors that were well-defined for both an edge and a region—common fate, blur similarity, color similarity, orientation similarity, proximity, and flicker synchrony—and found that all six factors showed figure-ground effects in the predicted direction, albeit to widely varying degrees.

Articulating motion—Barenholtz and Feldman (2006) demonstrated a dynamic principle of figure-ground organization: When a contour deforms dynamically, observers tend to assign figure and ground in such a way that the articulating (or hinging) vertices have negative (concave) curvature—which ensures that the figure side is perceived as containing rigid parts articulating about their part boundaries. In their experiments, this articulating-concavity bias was shown to override traditional static factors (such as convexity or symmetry) in cases where they made opposing predictions. In other experiments, Barenholtz (2010) showed that when a contour segment that is concave on one side and convex on the other deforms dynamically, observers tend to assign the figure on the convex rather than the concave side.

Advancing region motion—Barenholtz and Tarr (2009) showed that moving a border within a delimited space such that the bounded area on one side grows larger and the bounded area on the other side shrinks in size causes the growing area to be perceived as a figure advancing onto the shrinking area. Thus, motion in an advancing region overpowers the classic Gestalt factor of small area.

Contour entropy as a determinant of ground or hole—All of the above research is aimed at finding factors that determine the perception of a figure against a background. Gillam and Grove (2011) recently asked whether there are also factors that strengthen the perception of a region as ground. They reasoned that occlusion by a nearer surface will usually introduce a regularity among terminations of contours at the occluding edge, which will be perceived as a stronger cue to occlusion when the irregularity of the elements is higher. In other words, when the lines being terminated are more disordered, the strength of the occlusion cue (called “order within disorder” or “entropy contrast”) is larger. They predicted that unrelated (high entropy) lines would tend to appear as ground (or “holes”) in a figure-ground paradigm more often than more ordered (low entropy) lines, which was confirmed in three experiments. This work significantly expands earlier work on the perception of holes (e.g., Bertamini, 2006; Bertamini & Hulleman, 2006; Nelson & Palmer, 2001).

5.4 Nonimage-Based Influences on Figure-Ground Perception

5.4.1 Past experience—For most of the 20th century, the Gestalt view that past experience did not influence initial figure-ground perception prevailed. Part of the reason for this continued view might have been that few direct tests existed of the influence of past experience on figure-ground perception that satisfied Wertheimer’s (1923) criteria mentioned above. Either there was no evidence that past experience alone drove the effect, or it was possible that past experience exerted its influence after the initial perceptual reorganization. Other observer-dependent factors such as perceptual set and attention were also assumed to be high-level influences, operating only after figures had been segregated from grounds. Recently, evidence that past experience can influence figure assignment has accumulated both from direct report and indirect, response time measures.

Direct reports: Using direct report measures, Peterson, Harvey, and Weidenbacher (1991; Peterson & Gibson, 1994a) found that regions that portrayed portions of familiar objects were more likely to be perceived as figures when they were upright (and hence, portrayed the familiar object in its canonical orientation) than when they were inverted (see Figure 18). Importantly, regions portraying upright, but not inverted, familiar objects were more likely to be *obtained* as figures by reversing out of the alternative interpretation, suggesting that past experience in the form of familiar configuration exerts an influence on the initial determination of figure and ground (Gibson & Peterson, 1994; Peterson et al., 1991; Peterson & Gibson, 1994a, 1994b).

The stimuli used in these studies most likely satisfied Wertheimer’s first criterion for evidence of past experience effects on figure assignment. However, it was—and still is—difficult for experiments that measure subjects’ phenomenal reports to satisfy his second criterion (to exclude that past experience exerts its influence after some initial perceptual reorganization). Even with masked exposures as short as 28 ms, such as those used by Peterson and Gibson (1994a), it is still possible that some preliminary, unconscious, figure-ground determination preceded the organization about which observers reported (Epstein & DeShazo, 1961), introducing the need for an indirect or implicit measure.

Indirect measures: Driver and Baylis (1996) were the first to use reaction time measures to index figure-ground perception in an experiment in which subjects were shown a brief exposure of a display containing a small high contrast region lying on a larger rectangle. In this stimulus configuration, both contrast and the configural factor of small area biased participants to see the small region as figure (Figure 19A). On each trial, participants were asked which of two small, enclosed test shapes shown after the initial display had the same border as the stepped border dividing the rectangle into two regions. On half of the test trials, the test shape lay on the same side of the border as the region that was likely to have been perceived as figure (“figure” probes; Figure 19B), whereas on the remainder of the test trials, the test shape lay on the opposite side (“ground” probes; Figure 19C). Faster and more accurate responses were obtained on figure test trials than on ground test trials, suggesting that the border shared by two regions in the initial display was automatically bound to the one cued as figure. (For later use of Driver and Baylis’ indirect measure, see Hulleman & Humphreys, 2004; Vecera, Flevaris, & Filapek, 2004.)

Driver and Baylis (1996) also interpreted the results they obtained with their indirect measure to indicate that past experience does not influence figure assignment, reasoning that response times should be equally fast for figure and ground probes if past experience were operating. Peterson and Enns (2005), however, argued that this reasoning was flawed and that a past experience account could also predict the pattern of results obtained on ground test trials by Driver and Baylis (see also Peterson & Lampignano, 2003). This is because the repetition of the border from the first trial would reinstantiate the memory of where the figure lay when the border was first encountered, and this memory would compete with the current cues favoring the figure on the opposite side of the border. This competition would slow responses on ground test trials compared to figure test trials. Peterson and Enns introduced control test trials to investigate this alternative interpretation. They asked subjects to ignore the first display and then to decide for a second display, whether the pair of small closed figures it contained was identical (as a consequence, the first display served as a priming display). Consistent with the past experience account, response times were slower on ground test trials (and faster for figure test trials) compared to control trials facing in the same direction.

Further evidence for the role of past experience in figure-ground perception was obtained by Vecera and Farah (1997), who presented observers with overlapping outline letters and asked them to determine whether two probed locations were on the same letter or on different letters, a distinction that depended on figure-ground organization. Congruent with an influence of past experience, faster and more accurate responses were obtained when the outline letters were in their familiar upright orientation rather than inverted. Peterson and Skow (2008) went one step further using displays like those in Figure 18E-G, in which shape properties such as small area, enclosure, and symmetry strongly favored perceiving the figure on the inside of a silhouette’s borders, but past experience favored seeing the figure on the outside. The stronger factors favoring the figure on the inside dominated the competition for figural assignment. Nevertheless, responses to the familiar object that lost the competition were suppressed for a brief period of time after figure-ground perception was achieved, providing evidence that past experience plays a role in figure-ground segregation even when it does not dominate (see also Trujillo, Allen, Schnyer, & Peterson, 2010, for a study measuring the associated event related potentials). Recently, Navon (2011) conducted a series of experiments that he took as evidence that the familiarity of objects may exert an influence even earlier, when pre-figure assignment parsing occurs.

Taken together, these findings from studies that satisfy both of Wertheimer’s (1923) criteria show that past experience can exert an influence on several aspects of figure-ground perception.

5.4.2 Attention and perceptual set—Before any behavioral evidence was obtained that attention affects which region will be perceived as figure, Kienker, Sejnowski, Hinton, and Schumacher (1986; Sejnowski & Hinton, 1987) modeled figure-ground segregation as cross-border inhibitory competition that could be biased by attention. A decade later, Baylis and Driver (1995) presented empirical evidence for the role of attention. They found that when observers allocated their attention endogenously to one of two regions sharing a border, they were more likely to perceive the attended region as figure than the unattended region. Similar, but weaker effects of exogenous attention were later obtained (Vecera et al., 2004). In addition, it has been shown that the viewer's intention, or perceptual set, to perceive one of two regions sharing a border as figure can influence figure assignment (Peterson et al., 1991; Peterson & Gibson, 1994b). Thus, there is ample evidence that observer-dependent factors can influence figure assignment.

The effects of attention and perceptual set contradict claims that figure-ground perception *always* occurs “early” in the visual system. Indeed, although figure-ground perception can be affected by focused attention, there is evidence that it can *also* occur preattentively. Preattentive processes are considered “early” in that they are carried out before attention is focused on individual regions in the visual field, that is, when attention is either distributed broadly or allocated to a different task. For example, in a case study of a patient with left neglect (a failure to attend to the left side of objects and spaces), it was found that judgments of the symmetry of individual shapes in the left hemisphere were severely impaired (Driver et al., 1992). Nevertheless, when asked to make figure-ground judgments regarding six-region displays with green and red alternating symmetric and asymmetric regions, the patient, like controls, showed a bias to see symmetric regions as figures, suggesting that the patient could use symmetry preattentively for figure-ground segregation even though he could not detect it postattentively. Convexity-based figure-ground segregation can also occur preattentively. Kimchi and Peterson (2008) asked participants to perform a demanding change detection task on a stimulus presented on a task-irrelevant background of alternating regions of figures and grounds by convexity. They found that changes in the organization of the background produced congruency effects on target-change judgments, even though participants were unable to report the figure-ground status of the background. Taken together, these findings are consistent with the claim that figure-ground segregation can occur before focal attention is directed to making a figure-ground judgment. Thus, figure-ground segregation can occur preattentively, but it can also be affected by attention.

5.5 Figure-Ground Organization in Relation to Shape and Depth Perception

Traditionally, figure-ground organization has been studied as a relatively isolated, albeit important aspect of perceptual organization. Studies like those reviewed above, however, have clearly shown that figure-ground organization is closely related to other aspects of visual perception like shape and depth perception.

5.5.1 Shape perception—Consider the following demonstration due to Attneave (1971): A random wiggly contour is drawn across a disk, dividing it into two, and the two half-disks are spatially separated (Figure 20). Even though the contour is—by construction—identical on the two halves, it *looks* different. The wiggly contour on the left half-disk appears to contain three wide, smooth parts, whereas on the right, it appears to contain four narrow, pointy parts. This demonstration suggests that the visual system assigns shape descriptions to surfaces, rather than to the bounding contours of those surfaces: Even when instructed to judge a contour in isolation, observers cannot help but be influenced by the geometry of the surface to which the contour belongs. Hence, shape description is closely tied to figure-ground assignment.

One account of how these processes interact was offered by Hoffman and Richards (1984). They proposed that the visual system segments objects into parts at points of negative minima of curvature (points of locally highest curvature in concave regions) along the bounding contour of the shape. According to this minima rule, the same wiggly contour in Figure 20 is segmented differently on the two half-disks because, when figure and ground switch, so do the roles of convexities and concavities. Hence perceptual part boundaries—which lie in concavities—are located at different points on the two halves (marked with the dots in Figure 20). The contour is perceptually segmented into parts differently on the two half-disks—hence it looks different. This account of part segmentation has been supported in a number of experiments (e.g., Barenholtz & Feldman, 2006; Baylis & Driver, 1994; Cohen & Singh, 2007; De Winter & Wagemans, 2006; Singh & Hoffman, 2001).

Building on this account, Hoffman and Singh (1997) proposed that the relative salience of the two sets of parts, one at each side of the borderline, plays an important role in determining figure and ground. (In this respect, it could be considered one of the new image-based principles of figure-ground organization reviewed before.) The perceptual salience of a part, or part salience for short—how much it visually stands out as a separate part—is determined by a number of geometric factors, including the curvature (“sharpness”) of its part boundaries, and its protrusion (the degree to which it “sticks out,” measured as perimeter/cut length). Typically, the side with the more salient parts is assigned figural status, as illustrated in Figure 21. In this figure, (a) shows the original contour, (b) shows the parts obtained from the two sides, and (c) shows the side with more salient parts, which tends to be perceived as figure. Because part salience is based on relatively local computations, figure and ground can perceptually reverse along a continuous contour (see Figure 21D). This suggests that shape analysis occurs not only on the figure side—*after* figure-ground assignment has taken place—but actually precedes it, and it plays a crucial role in making the figure-ground assignment. This account is consistent with previous proposals made by Peterson and colleagues discussed before, although it involves only low-level shape analysis, and does not posit access to memories of familiar shapes.

Another way of thinking about this shape analysis is in terms of a skeletal or axial representation of shape. (Hence, axiality could also be added to the list of new image-based principles of figure-ground organization.) Skeletal or axial representations provide a compact “stick figure” representation that reflects the qualitative branching structure of a shape (Blum, 1973; Leyton, 1989; Marr & Nishihara, 1978). In a recent probabilistic model of shape skeletons (Feldman & Singh, 2006), the “goodness” of a skeleton—formally, its posterior $\text{pr}(\text{Skeleton}|\text{Shape})$ —is evaluated in terms of (i) how well the skeleton explains the shape—its likelihood $\text{pr}(\text{Shape}|\text{Skeleton})$; and (ii) its simplicity—the prior $\text{pr}(\text{Skeleton})$ which embodies a preference for fewer branches and less curved branches. The “best” skeleton (that maximizes the posterior) provides an intuitive summary description of the shape, and allows a one-to-one correspondence between its skeletal branches and the parts of the shape. The skeletal model of shape description applies to the figure-ground context as follows: The skeleton is computed from both sides of the contour, yielding skeletal descriptions of the surface on either side; and whichever side has a more “axial” shape (higher posterior) wins the figure-ground competition. A recent Bayesian implementation of this competition has been shown to closely resemble human perception of figure and ground (Froyen et al., 2010).

5.5.2 Depth perception—Because the figure is by definition closer to the observer than the ground, figure-ground organization is necessarily related to certain aspects of depth perception. One can think of figure-ground principles as a subset of sources of depth information that apply to the special case of inferring relative (i.e., ordinal) depth across an edge. Recently, the seemingly ordinal figure-ground factor of convexity was found to

combine with classic metric depth cues, such as binocular disparity (Burge, Peterson & Palmer, 2005), suggesting that it might be a metric rather than an ordinal depth cue. Indeed, in a recent study of natural scene statistics, Burge, Fowlkes, and Banks (2010) concluded that convexity is a metric cue to depth. Many questions remain regarding how figural cues signal depth.

5.6 Conclusion

Research on figure-ground organization has also come a long way since Wertheimer's (1923) paper on the Gestalt laws. First, when studied in more controlled experiments, the classic, configural principles turn out weaker than previously supposed. Second, using richer displays, new figure-ground principles have been discovered with a plausible ecological foundation. Third, the role of past experience and attention has now been clearly demonstrated in experiments that satisfy Wertheimer's own criteria. Finally, figure-ground organization is no longer studied in isolation but turns out to be intimately related other processes such as shape and depth perception. These connections are currently also the focus of interesting research on the neural mechanisms of contour grouping and figure-ground organization, as reviewed next.

6 Neural Mechanisms in Contour Grouping, Figure-Ground Organization, and Border-Ownership Assignment

6.1 Introduction

Recall that the failure of the experiments by Lashley et al. (1951) and Sperry et al. (1955) to support key predictions of Köhler's electromagnetic field theory of brain function was a devastating blow to Gestalt theory's attempt towards a radically different understanding of the brain mechanisms that underlie perception and cognition. The dominant model of brain function since then has been that the brain consists of an architecture of organized, interconnected neurons that interact via synaptic communication to form complex circuits that respond selectively to different properties of visual stimulation. Building on Hubel and Wiesel (1968)'s discovery and characterization of four types of neurons in the primary visual cortex (concentric, simple, complex and end-stopped), and the large number of neurophysiological studies that followed since then, the current view of the visual system is that of a hierarchy of stages of gradually increasing scale and complexity of processing, paralleled by the increasing receptive field size and increasing selectivity of neurons across the sequence of cortical areas in the "ventral stream" (see Rust & DiCarlo, 2010, for a quantitative study).

It may seem that such a conception is fundamentally incompatible with Gestalt notions of brain function, but this is untrue. First, there is some evidence against such a strict hierarchical organization. For instance, measurements of receptive field size tend to underestimate the influence of context in neurons at the low levels, as many studies of so-called non-classic surround influences have shown (Allman, Miezin, & McGuinness, 1985; for further references, see Alexander & van Leeuwen, 2010; Angelucci et al., 2002). Second, and more fundamentally, there is a more abstract level at which a Gestalt conception of brain function might be correct even if Köhler's conjecture about its implementation in dynamically converging electromagnetic brain fields is not. It is now clear that recurrent networks of neuron-like elements—networks that contain closed feedback loops—are a much more plausible implementation of the hypothesis that the brain is a physical Gestalt. Physicist John Hopfield (1982) clarified this possibility when he proved that symmetric recurrent networks—networks with equal weightings in both directions between any pair of units—will always converge to an equilibrium state that satisfies an informational constraint isomorphic to minimum energy in physics. Hence, the Gestaltists may have been wrong in

detail about the brain being an electromagnetic Gestalt system, but they may still have been correct at a more abstract level if recurrent networks are found to be crucial for perceptual organization. Important models of organizational phenomena within recurrent networks have indeed been proposed by Grossberg and others (e.g., Grossberg & Mingolla, 1985; Kienker et al., 1986). Furthermore, while neurophysiological models for contour integration based upon good continuation principles have been based primarily upon cortical networks in area V1 (Li, 1998; Roelfsema, 2006; Yen & Finkel, 1998), as well as recurrent interactions between areas V1 and V2 (Neumann & Sepp, 1999), fMRI studies of contour grouping in both human and macaque implicate not only V1 and V2 but other extrastriate visual areas such as V4 and LOC (e.g., Altmann, Bühlhoff, & Kourtzi, 2003; Kourtzi, Tolias, Altmann, Augath, & Logothetis, 2003).

Rather than providing a list of putative neural correlates of all the Gestalt principles, this section reviews neurophysiological studies investigating the neural mechanisms in contour grouping, figure-ground organization, and border-ownership assignment in an integrated way. In doing so, we demonstrate how contemporary neuroscience has embraced Gestalt ideas, while doing justice to Hubel and Wiesel's heritage in the following three ways: (1) We demonstrate how the responses of cortical neurons can depend on the parameters of the stimulus in its receptive field as well as on the properties of the overall configuration in the visual field; (2) we substantiate the Gestalt postulate of autonomous organization processes that form primary units of perception; and (3) we refine our understanding about the role of attention in these processes of perceptual organization.