VISUAL COGNITION, 2006, 14 (4/5/6/7/8), 619–628

Psychology Press
Taylor & Francis Group

Colour as a Gestalt: Pop out with basic features and with conjunctions

James R. Pomerantz
Rice University, Houston, TX, USA

Gestalt phenomena are a cornerstone of perceptual psychology. Although sometimes poorly understood, they are powerful and robust effects with significant implications for how we recognize objects and parse scenes. Traditionally, the study of Gestalts has focused on visual form perception, where parts combine in nonadditive ways to create wholes possessing novel emergent properties different from the "sum of their parts". Here I argue that colour perception meets the customary criteria applied to Gestalts at least well as shape perception does, in that colour emerges from nonadditive combination of wavelengths in the perceptual system and results in novel, emergent features. Regarding colour as a (and perhaps as *the* quintessential) Gestalt may help demystify Gestalts and help us better understand the role of colour in tasks such as visual search that are used to identify basic features in early vision. Colour should be thought of not as a basic feature or primitive property of the stimulus but rather as a complex conjunction of wavelengths that are integrated in perceptual processing. As a Gestalt, however, colour serves as a psychological primitive and so, as with Gestalts in form perception, it may lead to pop-out in visual search. Indeed, pop-out should be regarded as a prerequisite for claiming that a conjunction of features forms a Gestalt.

Students of psychology learn early on about Gestalts, beginning with the ubiquitous series of illustrations appearing in the early chapters of introductory textbooks. There we read attempts to translate this elusive German word into scientific terms, or at least into other languages, and we witness a spate of demonstrations showing various Gestalt laws of grouping and figure–ground segregation at work. As we read more deeply into the literature, including current research on Gestalt phenomena, we learn that despite significant progress towards understanding and operationalizing the concept (Beck, 1982; Kimchi, Behrmann, & Olson, 2003; Palmer, 1999; Pomerantz & Kubovy, 1981), in many respects it remains vague.

Please address all correspondence to James R. Pomerantz, Department of Psychology MS 25, PO Box 1982, Rice University, Houston, TX 77251-1892 USA. Email: pomeran@rice.edu

http://www.psypress.com/viscog DOI: 10.1080/13506280500195052

WHAT IS A GESTALT?

Paraphrasing William James's famous pronouncement on attention, we might say that everybody knows what a Gestalt is. At least we all know one when we see one. Gestalts, most fundamentally, are patterns or higher order features that emerge when two or more perceptual elements are placed in close spatial or temporal proximity to one another, patterns or features that do not arise when only a single element is present. So when three disks are tossed onto a surface, one perceives not only the properties of each disk individually (its size, colour, and x-y coordinates) but also the configuration the three disks form, be it a triangle or a straight line. If each disk is shifted in the same direction and by the same distance, the configuration remains unchanged even though the x-y coordinates of all the disks have changed. In that sense, the "whole" created by the disks—the configuration or shape they form—is different from and independent of its parts. Should the three disks contain suitably placed notches, as in Kanizsa's (1979) well-known subjective contour figure, yet other Gestalts emerge. Likewise, should the three disks be enclosed within a circle, a face-like configuration might emerge, with individual disks serving as eyes or mouth.

GESTALTS OUTSIDE OF FORM PERCEPTION

Although Gestalts are not confined to visual form perception, that is where most examples are drawn and where most research has focused. We know, however, that Gestalts can arise in auditory perception, as when three tones played in sequence define a melody that remains invariant if each tone's frequency is changed proportionally or, when played simultaneously, define a chord. Apparently absent from the literature, however, is the notion that colours may be regarded as Gestalts, perhaps as quintessential Gestalts.

Of course colour does figure into the demonstration of other Gestalt principles, as when disks in Wertheimer's (1923, translated in part in Ellis, 1950) grids group into rows or columns based on colour similarity; or when colours divide into separate translucent layers following Metelli's (1974) scission phenomenon. Missing from mention, however, is the notion that colour itself is a Gestalt, an emergent phenomenon arising from the nonadditive combination of more primitive components. Upon reflection, I believe colour meets most, if not all, of the criteria that customarily define a Gestalt, and that by regarding colour as a Gestalt we may sharpen our yet-vague concept and perhaps peel away some of the mystery and confusion that always seems to bedevil discussions of Gestalts.

COLOUR VISION AS GESTALT PROCESSING

Although controversies remain in our understanding of human trichromatic vision, much of it is well understood. In particular, we know that at the earliest stages of coding, three separate receptor systems exist with overlapping spectral sensitivities—cones with overlapping tuning curves indicating their response to a range of wavelengths. At later stages of processing, these separate channels for short, medium, and long wavelengths are integrated via an opponent process network into signals representing a red–green axis, a yellow–blue axis, and a black–white (luminance) axis. Yet later stages of processing convert this representation into one that codes hue, brightness, and saturation (which are taken to be the three dimensions of conscious perceptual experience).

Does the recoding that takes place in colour vision resemble that which we observe with the perception of visual forms, with the whole being different from—having novel, emergent properties apart from—the sum of its parts? The nonlinear way in which wavelengths combine to produce subjectively experienced colours certainly seems Gestalt in nature. After all, not only do wavelengths mix to form colours quite different from the colours associated with the individual wavelengths alone, but those emergent colours do not appear as simple blends and often they are unpredictable, at least to perceivers not well-versed in colour vision theory. Many of the colours arising from wavelength mixtures are in no way similar to the colours perceived from their constituent wavelengths, and in fact these emergent colours often do not appear in the visible spectrum. A good example is white itself, which does not appear in the spectrum of a trichromat and could be viewed as every bit as much of a novel, emergent, and surprising feature as Kanizsa's (1979) subjective triangle. What is more, just as there is an infinite number of tonal sequences that can define the same melody, there is an infinite number of wavelength combinations that yield the perception of white.

WHOLES BEFORE PARTS

If colour can be viewed as a Gestalt, might an even stronger case be made that colour is *the* quintessential Gestalt? I believe that it can, for the following reason. One principal claim of Gestalt psychology is that, in perception, the whole precedes the parts. Although this argument can become complicated and has been framed in various ways, the central thesis is that the forest is seen before the trees, i.e., that global structures or wholes are perceived via a fast, primary ("preattentive") process, whereas local structures or parts are perceived only by a slow, secondary process that relies on focal attention. The evidence for this notion of global-first in form perception is mixed; in some cases it seems to apply (Navon, 1977), whereas in others it holds for only some of the time or not at all (Kinchla & Wolfe, 1979; Pomerantz, 1983; Pomerantz & Sager, 1975; for a review see Kimchi, 1992). Much depends on physical parameters of the stimulus such as spatial frequency and absolute size: If one looks at a newspaper photograph from a distance, one sees only the global scene and not the dots from which that scene emerges. If the stimulus is brought close enough to the eye, however, the dots prevail.

Does holistic perception dominate in the case of colour? The answer appears to be yes. Human vision, in fact, seems incapable of attending selectively to individual wavelengths or of recovering the underlying spectra of visual stimuli, a fact well attested to by the existence of metamers (identically appearing colours that arise from different underlying wavelength combinations). Unlike in form perception, where usually we can attend either to the parts or to the whole (albeit not with equal ease or efficiency), in colour perception, the parts—the constituent wavelengths—are not accessible to conscious perception or available for behavioural responses.

MULTISTABILITY

It may be that for just this last reason—the inaccessibility of the component parts—that colour is not usually thought of as a Gestalt. Most Gestalt effects involve multistability—the flip-flopping that accompanies our inspection of a Necker Cube, the Rubin faces–vase stimulus, the Bahnsen columns, Wertheimer's lattices, and other classic demonstrations. With most Gestalt effects, one can switch attention between the parts and the wholes and sometimes even from one component part to another. Much of the "gee-whiz" factor that makes Gestalt effects so compelling arises from this metaperceptual experience (Pomerantz & Kubovy, 1981) wherein observers are made aware of the workings of their own perceptual processes. With colour, we do not experience this multistability. Instead, we see only the end product of our visual system's combining parts automatically, involuntarily, and irreversibly into the perceptual wholes we experience as colours. I would argue that colour should nonetheless be regarded as a Gestalt despite its lack of multistability: Just as with other stimuli (such as real-world scenes viewed binocularly), with colour one perceptual organization of wavelengths proves so strong it simply dominates any alternatives and there is then no perceptual flip-flopping.

THE EXPERIENCE ERROR

Conceptualizing colour as a Gestalt may aid us in our understanding of holistic perception in other areas. As is recognized by many (see Palmer, 2003), the concept of Gestalt is now, as it has always been, somewhat confused and vague. Indeed, the very discussion of critical issues such as object perception is clouded by a lack of an operational definition of "object" (e.g., Wolfe & Bennett, 1997). When we conduct experiments and describe our stimuli, we often use terms like "stimulus", "item", "object", and so forth when, in truth, we would be hard pressed to defend ourselves in the face of claims that we are committing the "experience error".

In writing the methods section of a scientific paper on perception, if we refer to a stimulus as a "red" object, we are committing the experience error. This error was a concern for the Gestalt psychologists but has been widely ignored by contemporary researchers (save for Palmer, 1999). It was described by Wolfgang Köhler (1929/1947, p. 95) as follows:

> In psychology we have often been warned against the stimulus error, i.e., against the danger of confusing our knowledge of the physical conditions of sensory experience with this experience as such. As I see it, another mistake, which I propose to call the experience error, is just as unfortunate. This error occurs when certain characteristics of sensory experience are inadvertently attributed to the mosaic of stimuli.

In brief, we commit the experience error when we attribute a phenomenon that is a product of cognition—i.e., something in the head—to the stimulus—i.e., something in the outside world. Even though he knew nothing of the electromagnetic spectrum, Newton nonetheless knew that colour is not in the stimulus: "The rays, to speak properly, are not coloured. In them there is nothing else than a certain Power and Disposition to stir up a Sensation of this or that Colour" (Newton, 1704/1952, pp. 124–125). After all, radio waves, microwaves, and gamma rays lack colour to human eyes, so one might speculate what we might say in reply to an alien species that used colour terms for those portions of the total spectrum.

Instead of saying that a stimulus is red, we should say that is has a particular spectrum that is seen by human trichromats as red. This substitution would convey two benefits. The first is precision: To call a stimulus "red" vastly underspecifies the stimulus, given that there is potentially an infinite number of wavelength combinations that can produce the identical apparent shade of red. The second is explanatory power: To ascribe a perceptual outcome to the colour red is to explain one perceptual experience (or behaviour) in terms of another, and such perceptual causality quickly gets complex and hard to justify. The goal of cognitive science and psychophysics is, ultimately, to explain perception in terms of the stimulus, even though it may be—and often is the case—that cognitive mediation accounts for much of what we see.

IMPLICATIONS FOR PERCEPTUAL EXPERIMENTATION: VISUAL SEARCH

There are real and important implications of the experience error for ongoing perceptual research. Consider the burgeoning area of visual search, a field that holds the promise for understanding both the basic features underlying early vision and the features that guide later, attention-demanding visual processing (Treisman & Gelade, 1980; Wolfe, 1994; Wolfe & Horowitz, 2004).

One of the main tools in visual search studies is the calculated slope of the function relating reaction time (RT) to the number of items in a display to be searched. When searching for a red disk in a field of green disks, RTs to indicate whether the red target is present are largely independent of the number of green disks displayed. This indicates that colour is probably a basic feature, one that pops out of a display without having to be searched for sequentially. Targets defined by features such as colour, size, orientation, movement, and the like often pop out and so are thought to be primitives on which early vision is founded.

Targets defined by conjunctions of primitive features, by contrast, usually do not pop out and so must be detected through slow, inefficient search. Thus, searching for a red disk in a field of red squares and green disks is slow, presumably because it takes the second, attention-demanding step of feature integration to identify the target.

Thinking back to what has been noted above about colour and Gestalts, we must be careful in describing our stimuli to avoid committing the experience error. In calculating search slopes, for example, we divide RTs by the number of items in a display. In doing so, however, we must be sure we understand that number accurately, and that can prove tricky because, as noted above, we have no accepted definition for an "item" any more so than we have for an object, a stimulus, or even a "thing". If a display contains 32 disks that are spaced into pairs, do we say that the display contains 32 disks or 16 disk pairs?

Returning to colour, we must again consider that this feature resides in our heads and not in the stimulus. We should talk properly of wavelength combinations that lead to pop out, rather than speak of colours that pop out. The fact that we do not do so routinely, however, might not be so serious a matter. Talking about spectra can be cumbersome, after all, so in using the term "colour" might we simply be engaging in a universally understood shorthand?

Perhaps not. The point about colour, in terms of theories of visual search, is that basic features such as colour usually pop out, whereas conjunctions usually require focal attention. However, we must recall that colour itself is a conjunction of wavelengths! Thus, by taking our shorthand too literally, we run the risk of building contradictions into our theories.

If it is best to think of colour as a conjunction of wavelengths (or, taking it one step further into the language of sensory transduction, as a conjunction of the outputs of our short, medium, and long wavelength cone systems), would colour stand as the single exception to the principle that conjunctions do not lead to pop out? No, in part because there are other conjunctions that are detected quickly, such as Wolfe and Horowitz's (2004) example in which a black X will pop out from a field of white Xs and black Os. More directly relevant to the current argument, however, there are other instances of conjunction pop out that involve Gestalts. Consider, for example, the triple conjunction of the horizontal, vertical, and diagonal line segments that create arrows and triangles. As is shown in Figure 1, an arrow will pop out of a field of triangles as quickly as a black square pops out from a white background, even though the distinctive feature that makes an arrow differ from a triangle—an oppositely sloped diagonal line—often will not pop out from its mate (Pomerantz et al., 1977).

CONCLUSIONS

The primary conclusion to be drawn from these observations is that Gestalts are all about the integration of primitive elements into new units (or emergent features) that then become primitives in perceptual processing.

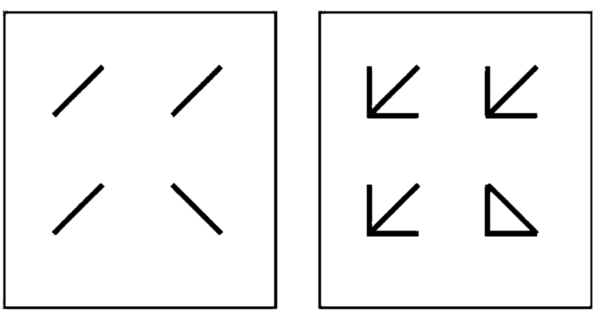

Figure 1. In the left panel, RTs to locate the oddly sloped diagonal are long, indicating that it fails to pop out. In the right panel, identical L-shapes have been added to each diagonal, and now the triangle pops out from the arrows as quickly as a black square pops out from a white background (adapted with modifications from Pomerantz, Sager, & Stoever, 1977).

Unlike the integration of separable features that do not yield Gestalts (an integration that is slow and attention demanding), the integration of elements that yield Gestalts is fast and seemingly automatic in the sense of requiring no cognitive resources and being outside the reach of voluntary control.

In the case of wavelengths, the integration into perceived colours is also seemingly irreversible, with wavelength-specific information flowing through "sealed channels" (Pomerantz, 1978, 2003) that are not accessible to consciousness or available for response purposes. Functionally speaking, the whole of a colour such as white precedes the parts, i.e., the constituent wavelengths from which colour arises (cf. the reverse hierarchy notion of Hochstein & Ahissar, 2002; see also DiLollo, Kawahara, Zuvic, & Visser, 2001). So fast and complete is the perceptual integration of wavelengths that we are unaware that our visual system has created a Gestalt.

Should this line of reasoning prove correct, it suggests that whenever parts conjoin to form Gestalts, those conjunctions should give rise to very low or zero search slopes in RT experiments. There is good evidence of this from various configural superiority effects reported previously (Pomerantz et al., 1977). In fact, in some of those experiments where display elements group to create emergent features, doubling the number of homogeneous distractors drastically *reduced* RTs to localize the singleton target; see Figure 2. The result was *negative* search slopes, steeper than −200 ms/item. This illustrates again how powerful Gestalt contextual effects can be and how they can be measured using standard tools of cognitive psychology.

In any case, the strong argument here is that Gestalts must yield pop out if we are to regard them as proper Gestalts. Any jumble or conjunction of features could be proposed as a Gestalt, but for that claim to be credible, that conjunction should show pop out. In turn, this new and explicit criterion will help us clarify what is and is not a Gestalt.

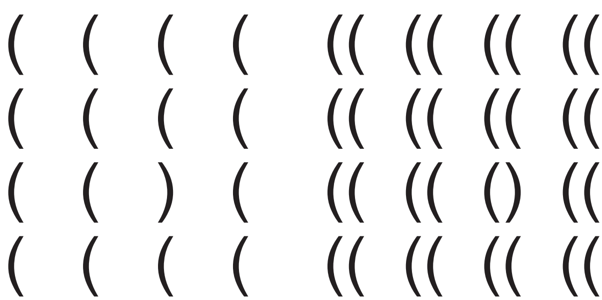

Figure 2. Locating the rightward curving line segment takes longer in the left panel, where there are only 15 leftward curving distractors, than in the right panel, where there are 31 distractors (adapted with modifications from Pomerantz et al., 1977.) Under some conditions, the RT difference results in steeply negative search slopes exceeding −200 ms/line segment.

Gestalt phenomena are fascinating and important, but they have proven frustrating to study because of the difficulty of operationalizing them, in developing metrics for them, and in modelling them. By comparison with form perception, our understanding of colour perception is detailed and sophisticated. As Palmer (1999) has pointed out, colour can be seen as a useful microcosm of the entire perceptual system. If we regard colour as a Gestalt and approach the integration of elements of form into perceived shapes in ways analogous to how we model sensory integration of wavelength-specific information in the colour domain, we may strip away some of the vagueness that has accompanied Gestalt psychology and make significant progress towards understanding the perception of visual wholes.

REFERENCES

Beck, J. (1982). *Organization and representation in perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

DiLollo, V., Kawahara, J., Zuvic, S. M., & Visser, T. A. W. (2001). The preattentive emperor has no clothes: A dynamic redressing. *Journal of Experimental Psychology: General*, *130*, 479–492.

Ellis. W. D. (Ed.). (1950). *A sourcebook of Gestalt psychology*. New York: Humanities Press.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804.

Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, *112*, 24–38.

Kimchi, R., Behrmann, M., & Olson, C. R. (2003). *Perceptual organization in vision: Behavioral and neural perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kinchla, R. A., & Wolfe, J. M. (1979). The order of visual processing: Top-down, bottom-up, or middle-out. *Perception and Psychophysics*, *25*, 225–231.

Köhler, W. (1947). *Gestalt psychology*. New York: Liveright. (Original work published 1929)

Metelli, F. (1974). The perception of transparency. *Scientific American*, *230*, 90–98.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.

Newton, I. (1952). *Opticks, or a treatise of the reflections, refractions, inflections, and colours of light* (4th ed.). New York: Dover. (Original work published 1704)

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Palmer, S. E. (2003). Perceptual organization and grouping. In R. Kimchi, M. Behrmann, & C. R. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pomerantz, J. R. (1978). Are complex visual features derived from simple ones? In E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal theories of visual perception* (pp. 217–229). Chichester, UK: John Wiley.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General*, *112*, 516–540.

Pomerantz, J. R. (2003). Wholes, holes, and basic features in vision. *Trends in Cognitive Sciences*, *7*(11), 471–473.

Pomerantz, J. R., & Kubovy, M. (1981). Perceptual organization: An overview. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 423–456). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pomerantz, J. R., & Sager, L. C. (1975). Asymmetric integrality with dimensions of visual pattern. *Perception and Psychophysics*, *18*, 460–466.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their compound parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 422–435.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. *Psychologische Forschung*, *4*, 301–350.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202–238.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, *37*(1), 25–44.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 1–7.