VISUAL COGNITION, 2005, 12 (4), 639–690

Psychology Press
Taylor & Francis Group

Commentaries

Stability and change

James Todd

Department of Psychology, Ohio State University, Columbus, USA

Most people would agree that the shape of an object is one of its most perceptually important attributes, and some researchers have argued that it is the primary attribute by which observers are able to recognize objects (e.g., Biederman, 1987). Given the ubiquity of this common intuition, it is somewhat puzzling to note that the concept of ''shape'' has no formal mathematical definition that can adequately characterize its intended meaning when used colloquially. For example, almost everyone would concur that a big sphere and a small sphere both have the same shape, yet by most of the standard measures used in geometry they are quite different.

The abstract nature of the concept of shape is perhaps best revealed by the perceptual classification of biological forms (e.g., see Thompson, 1942). Consider, for example, the ability of normal individuals to identify their friends and loved ones under a variety of different viewing conditions. We are able to identify people from different vantage points, and with different facial expressions, hairstyles, make-up, or clothing accessories, such as hats or jewellery. We can also identify people after they undergo a growth spurt, gain or lose weight, or suffer the effects of ageing. These observations suggest that the identity of an individual's face must be based on some remarkably abstract property that is somehow unaffected by all of the transformations that faces typically undergo in the natural environment.

This core idea of invariance under change has been of great importance to the development of modern geometry. In 1872, the German mathematician Felix Klein gave a lecture at Erlangen University, in which he outlined a general principle for constructing different geometries that is now known as the Erlanger Program. His basic idea was to consider arbitrary groups of single valued transformations, and to investigate the properties of objects they leave invariant. Klein noted that some object properties are more stable than others, because they remain invariant over a larger set of possible transformations. By using different properties to define ''shape equivalence'', it is possible to create a hierarchy of

http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280444000265

geometries (i.e., Euclidean, affine, conformal, etc.) in which structural properties can be stratified with respect to their stability in a formally precise way.

Since the 1980s, Lin Chen has been developing a provocative hypothesis that a similar stratification of object properties may also exist in human perception, and he has accumulated a large body of evidence to support this view, which is elegantly summarized in the preceding target paper. His primary conclusion based on this evidence is that the ordering with which various properties are processed within the human visual system is based on their relative stability. Thus, topological properties are the quickest to be processed, whereas metrical properties are the slowest—precisely the opposite of what would be expected by many current researchers in the field.

I have been sympathetic to Chen's general approach since I first read his report in *Science* on this topic in 1982. At that time, my own research was focused on the perceptual identification of different categories of events. For example, human observers can easily distinguish between rigid and nonrigid motion (Todd, 1982). They can identify a wide variety of gaits in humans and other animals, such as walking, running, skipping, or dancing (Johansson, 1975). They can identify slow viscal-elastic changes in form, such as growth or weight gain (Todd, Mark, Shaw, & Pittenger, 1980), and they are also quite good at identifying intentional, social interactions, such as affection or aggression (Heider & Simmel, 1944). A remarkable property of event perception is that these different types of change can easily be identified even when they are applied to an unexpected structural configuration such as a random pattern of dots. The reason Chen's paper resonated with me in 1982 is that it was closely related to a hypothesis proposed by several other researchers concerning possible sources of information by which different transformations could potentially be identified (see Pittenger & Shaw, 1975; Todd et al., 1980). According to this approach, each type of event may be distinguished by the specific set of object properties it alters, and the specific set of properties it leaves invariant.

A fundamental fact of the natural environment is that all objects undergo change. Human observers possess the ability to identify a specific type of transformation (e.g., rotation) over a wide range of possible objects, and they are also able to identify a specific object (e.g., my wife) over a wide range of possible transformations. Within that context, it should not be surprising that early visual processing should focus primarily on those relatively stable properties that define the essence of an object's identity, rather than those that have a much higher variance under natural viewing conditions. There are many common events in our day-to-day experiences that highlight the importance of stability over change in object recognition. For example, I often fail to notice—at some cost—when my wife comes home with a new hairstyle, but I have never once mistaken her for someone else.

There is another important reason why the human visual system might allocate its resources disproportionately to attributes of objects that have a relatively high degree of stability. Although the objects encountered in natural vision are most often three-dimensional, the visual information by which they are specified is confined to the two-dimensional projection surface of the retina. This projective mapping produces an inevitable loss of information, such that any given pattern of optical stimulation may have an infinity of possible 3-D interpretations. Recent research has shown, however, that these ambiguities can be highly constrained in the sense that all of the possible interpretations are related by a limited class of transformations (see Koenderink, van Doorn, Kappers, & Todd, 2001). Attributes of objects that are invariant over those transformations can be determined quite easily from the available visual information, whereas other properties can only be estimated by incorporating prior knowledge or adopting some ad hoc heuristic. It makes sense, therefore, that the human visual system would concentrate its resources on object attributes that can be determined most reliably.

Although I am in general agreement Chen's overall argument, there is one caveat I would offer about accepting alternative geometries developed by mathematicians too literally as potential models of visual information processing. It is important to keep in mind that mathematics is an idealistic endeavour that demands perfect rigor and consistency. Biology, in contrast, is much more pragmatic. Simple heuristics can evolve quite readily if they impart some benefit to the survival of a species, even if they do not conform to the rigorous standards formal geometry. In Part V of his target paper, Chen presents a well-known figure from Minsky and Papert (1969) that was originally devised to show the computational difficulty of topological relations (see also Todd, Chen, & Norman, 1998). Chen argues that this apparent counterexample to his thesis is due to other psychological factors, rather than a general inability to perceive topological relations. I agree with this assessment. However, by imposing unspecified boundary conditions on the visual processing of topological structure, there is a danger that the ''topology'' of visual perception may turn out to be substantially different from the ''topology'' of formal mathematics.

REFERENCES

Biederman, I. (1987). Recognition-by-components: A theory of human image interpretation. *Psychological Review*, *94*, 115–147.

Chen, L. (1982). Topological structure in visual perception. *Science*, *218*, 699–700.

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 243–259.

Johansson, G. (1975). Visual motion perception. *Scientific American*, *232*, 76–88.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2001). Ambiguity and the ''Mental Eye'' in pictorial relief. *Perception*, *30*, 431–448.

Minsky, M. L., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry. Cambridge*, MA: MIT Press.

Pittenger, J. B., & Shaw, R. E. (1975). Aging faces as viscal elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 374–382.

Thompson, D. W. (1942). *On growth and form*. Cambridge, UK: Cambridge University Press.

Todd, J. T. (1982). Visual information about rigid and nonrigid motion: A geometric analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 238–251.

Todd, J. T., Chen, L., & Norman, J. F. (1998). On the relative salience of Euclidean, affine and topological structure for 3D form discrimination. *Perception*, *27*, 273–282.

Todd, J. T., Mark, L. S., Shaw, R. E., & Pittenger, J. B. (1980). The perception of human growth. *Scientific American*, *242*, 132–144.

Please address all correspondence to: James Todd, Department of Psychology, Ohio State University, 142 Townshend Hall, Columbus, OH 43210, USA. Email: Tod.44@osu.edu

A neural basis for global object features

Robert Desimone

NIH, Bethesda, MD, USA

In my commentary, I will focus on the relationship between Lin Chen's topological model for visual processing and the known biological properties of the visual system.

It is well-established that object recognition in primates is mediated by processing in a hierarchy of visual areas that form the occipitotemporal processing pathway, or ventral stream (see Ungerleider & Mishkin, 1982). Visual processing at the earliest stage (V1) is based on ''local'' (within small receptive fields) analysis of relatively simple visual features, and processing becomes increasingly more ''global'' (large receptive fields), more complex, and more attuned to global object properties as one proceeds along the pathway into the interior temporal (IT) cortex. Most psychological accounts of perception are consistent with the neurophysiological view. Typical psychological models specify that the beginning, or elementary, features of visual perception can be largely identified with the local features processed by neurons in V1. These local features are then combined across different feature domains in extrastriate cortex and then elaborated into representations of figures versus ground and, ultimately, into global representations of complex objects such as faces and houses. How can the neurophysiology and anatomy of perception then be reconciled with Chen's novel psychological model, in which visual processing proceeds largely from a global to local direction, i.e., beginning with the highest levels of object topology and proceeding to the lowest. According to the Chen model, local object features can only be extracted from these global representations, later in time, and with greater effort. Recent neurophysiological data suggest that the Chen model may indeed be correct, and that the explanation lies in critical functional differences between the feedforward and feedback flow of information along the occipitotemporal pathway.

The feedforward flow of information along the occipitotemporal pathway is hierarchical and causal, in the sense that damage to early components of the pathway can largely deafferent later components, completely depriving cells at higher levels of visual input (see Desimone & Ungerleider, 1989). Consistent with the anatomy and physiology, lesions of V1 cause blindness in primates (with the exception of modest ''blindsight'' functions), whereas lesions of areas further along the pathway impair only higher level object recognition, such as an inability to recognize colours, or faces, for example. The earliest neuronal firing latencies recorded in each area also follow the anatomical progression, with the shortest latencies found in area V1 and the longest latencies found in IT cortex (but with considerable overlap, e.g., Schmolesky et al., 1998) However, it would be a mistake to assume that neural activity at the lowest levels of the visual system directly influences perceptual awareness or perceptual judgements at the level of simple visual primitives. In the extreme, we know that this cannot be the case, since we are not directly aware of neural activity in the retina, and the feature primitives for object perception are not the unfiltered discrete points of stimulation on the photoreceptor surface of each eye. We are only aware of the activity of our receptors indirectly, through their ultimate impact on neurons at higher levels of the visual system, and, perhaps beyond the visual system entirely. Thus, the primitives of visual perception must be extracted at a higher level—but where?

Even beyond the retina, there are several reasons to believe that the activity of neurons in lower order cortical areas only influences visual processing and enters awareness by virtue of its impact on neurons in higher order areas. One line of reasoning is based on the anatomical connections of the visual cortex (see Adams, Hof, Gattass, Webster, & Ungerleider, 2000; Baizer, Desimone, & Ungerleider, 1993; Baizer, Ungerleider, & Desimone, 1991; Webster, Bachevalier, & Ungerleider, 1993; Weller, Steele, & Kaas, 2002). The direct, feedforward, anatomical projections of area V1 are to visual areas further along the ventral stream rather than to structures outside the visual system, such as prefrontal cortex or the amygdala, which are more likely sites for decision making and behavioural output. Although V1 projects to the thalamus, its projections to the LGN are only feedback projections, and its projections to the pulvinar are only to the parts that interconnect with prestriate areas such as V2 and V4, not to parts that project outside the visual system. Even area V2 has only modest projections to structures outside of the visual cortex and visual thalamus. It is only by the time one reaches area V4 along the ventral stream that one begins to find substantial projections to nonvisual structures, and these types of projections increase again, substantially, by area IT.

Neurophysiological studies also suggest that neural activity in V1, and probably other lower order areas of the ventral stream, only indirectly influences object recognition and awareness. In binocular rivalry paradigms in monkeys, for example, the activity of the large majority of V1 neurons does not correlate with the animal's reports of what it sees. Substantial correlations between neuronal responses and perceptual judgements in rivalry paradigms are found only in IT cortex (Sheinberg & Logothetis, 1997). Attentional studies also support this distinction. The majority of studies of attentional effects on V1 neurons in monkeys show little or no influence of attention, at least over short time intervals (e.g., Luck, Chelazzi, Hillyard, & Desimone, 1997; Marcus & van Essen, 2002). Thus, even when the monkey evidences little or no awareness of stimuli outside the focus of its attention, V1 neurons respond vigorously to these unattended stimuli. Brain imaging studies do tend to show attentional effects on the BOLD response in V1, but these effects are smaller than in extrastriate areas and may reflect feedback projections to V1 (Pessoa, Kastner, & Ungerleider, 2003). Attentional effects on neurons in areas such as area V4 are considerably stronger than in V1, but mainly when two or more stimuli compete within the same receptive field (Desimone & Duncan, 1995; Luck et al., 1997; Reynolds, Chelazzi, & Desimone, 1999). When only a single stimulus occupies a given cell's receptive field, attentional effects on neuronal responses in V4 are found mainly for stimuli of low contrast, or with great task difficulty (Reynolds, Pasternak, & Desimone, 2000; Spitzer, Desimone, & Moran, 1988). Thus, even in area V4, there are many circumstances when the monkey may attend exclusively to a single location in the visual field, yet V4 neurons with receptive fields at other locations will respond vigorously to the unattended stimuli, of which the monkey is presumably unaware. Only in IT cortex do neurons show strong attentional effects for stimuli spaced throughout the visual field (Moran & Desimone, 1985).

Thus, the evidence from a variety of anatomical and neurophysiological studies suggests that information only from the later stages of processing in ventral stream (most likely IT cortex) *directly* enters awareness and influences perceptual or behavioural judgements. Since IT neurons are often selective for complex features of objects, such as their shape, this is consistent with the idea that the ''primitives'' for object recognition are mainly the high-level, or global, object features processed in IT cortex rather than the simple local visual features processed in V1, an idea that is also compatible with Chen's model. Furthermore, since some IT neurons show invariance for transformations of object features, such as their location, size, and orientation, this is also compatible with Chen's proposal that the primitives for object recognition are invariant over many common transformations (see Desimone & Ungerleider, 1989). High-level feature analysis in extrastriate areas such as V4 and TEO is less well understood, but there is accumulating evidence that complex features such as curvature are represented in V4, and these features may also show a limited degree of invariance over transformation (Pasupathy & Conner, 2002). Thus, even if it is ultimately established that cortical areas posterior to IT, such as area V4, contribute directly to perceptual judgements, this would not necessarily be inconsistent with the idea that the primitive features of object recognition are, in fact, high-level topological features.

Although Chen's model posits that object features with the highest degree of topological invariance have the greatest and most immediate effects on perceptual judgements, the model allows that lower level object features can also be extracted and utilized in perceptual tasks, albeit with reduced speed. The neural mechanisms to support this ability are not yet clear, but recent evidence suggests that feedback projections to lower order visual areas in the ventral stream might play a role. In difficult, attention-demanding tasks, V1 neurons have been reported to show attentional modulation of responses but at considerably longer latency than the latency of the visual response to the stimulus (Roelfsema, Lamme, & Spekreijse, 1998), whereas attentional effects on extrastriate are typically reported to occur soon after the onset of the visual response (Desimone & Duncan, 1995; Luck et al., 1997). Unfortunately, attentional effects on neuronal responses have not yet been studied across multiple cortical areas under identical task conditions, and thus it is not certain that the attentional effects on V1 responses occur later than those in extrastriate cortex. However, a recent study combing fMRI with MEG and ERPs in humans suggests this is indeed the case (Noesselt et al., 2002). If so, it would be consistent with the idea that neuronal processing in V1 is recruited into perceptual representations and judgements when needed, by virtue of feedback projections from higher order areas. Some of these connections may be strengthened over the course of perceptual learning (Lee, Yang, Romero, & Mumford, 2002). Of course, even in such tasks, the information in V1 would still need to be filtered through higher order visual areas because V1 has minimal direct anatomical input into structures that could support perceptual judgements or produce a behavioural output. Similar arguments could apply to the LGN, which a recent imaging study suggests may also be under attentional control, presumably through late cortical feedback (O'Connor, Fukui, Pinsk, & Kastner, 2002).

Although I have speculated on the direct contributions of neural activity in lower order versus higher order visual to perceptual awareness and perceptual judgements, it is of course possible to design perceptual tasks where threshold performance is limited by the properties of any given component of the visual system, including those involved in early visual processing. One can easily design perceptual tasks, for example, where colour or spatial acuity is limited by the properties of the retinal receptors. As argued above, however, the properties of these early neural components may be task limiting, but their influence on perceptual judgements may still be indirect, through anatomical projections to higher order areas. For any given perceptual task, it is an open question where the critical computations take place. The data from Chen and colleagues suggests that for many tasks, including long-range apparent motion (e.g., Zhuo et al., 2003), the computations take place in areas that contain representations of invariant object features.

In sum, it would be a mistake to conclude from the hierarchical anatomical organization of the ventral processing stream that the primitive features for object recognition are necessarily the low-level features processed at the level of cortical areas such as V1. Information in the ventral stream flows both ‘‘upstream’’ and ‘‘downstream’’, allowing for both high- and low-level features to be extracted depending on the perceptual task. Chen’s global-to-local topological model for perception is consistent with this neurophysiological scheme as well as with the psychological evidence that object representations are the primitives for many behavioural tasks. Chen’s model challenges neurophysiologists to take a new approach to feature analysis in the visual system, one that emphasizes object features that are invariant over transformations.

REFERENCES

Adams, M. M., Hof, P.R., Gattass, R., Webster, M. J., & Ungerleider, L. G. (2000). Visual cortical projections and chemoarchitecture of macaque monkey pulvinar. *Journal of Comparative Neurology*, *419*, 377–393.

Baizer, J., Desimone, R., & Ungerleider, L. G. (1993). Comparison of subcortical connections of inferior temporal and posterior parietal cortex in monkeys. *Visual Neuroscience*, *10*, 59–72.

Baizer, J., Ungerleider, L., & Desimone, R. (1991). Organization of visual inputs to posterior parietal and inferior temporal cortex in the macaque. *Journal of Neuroscience*, *11*, 168–190.

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In E. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. II, pp. 267–299). Amsterdam: Elsevier.

Lee, T. S., Yang, C. F., Romero, R. D., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature Neuroscience, 5*, 589–597.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2 and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*, 24–42.

Marcus, D. S., & van Essen, D. C. (2002). Scene segmentation and attention in primate cortical areas V1 and V2. *Journal of Neurophysiology*, *88*, 2648–2658.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in extrastriate cortex. *Science*, *229*, 782–784.

Noesselt, T., Hillyard, S. A., Woldorff, M. G., Schoenfeld, A., Hagner, T., Jancke, L., Tempelmann, C., Hinrichs, H., & Heinze, H. J. (2002). Delayed striate cortical activation during spatial attention. *Neuron*, *35*, 575–587.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*, 1203–1209.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*, 1332–1338.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2003). Neuroimaging studies of attention: From modulation of sensory processing to top-down control. *Journal of Neuroscience*, *23*, 3990–3998.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*, 1736–1753.

Reynolds, J. H., Pasternak, R., & Desimone, R (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*, 703–714.

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*, 376–381.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, *79*, 3272–3278.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceeedings of the National Academy of Sciences*, *94*, 3408–3413.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, *240*, 338–340.

Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Webster, M. J., Bachevalier, J., & Ungerleider, L G. (1993). Subcortical connections of inferior temporal areas TE and TEO in macaque monkeys. *Journal of Comparative Neurol*ogy, *335*, 73–91.

Weller, R. E., Steele, G. E., & Kaas, J. H. (2002). Pulvinar and other subcortical connections of dorsolateral visual cortex in monkeys. *Journal of Comparative Neurology*, *50*, 215–240.

Zhuo, Y., Zhou, T. G., Rao, H. Y., Wang, J. J., Meng, M., Chen, M., Zhou, C., & Chen, L. (2003). Contributions of the visual ventral pathway to long-range apparent motion. *Science*, *299*, 417–420.

Please address all correspondence to: Robert Desimone, NIH, Building 10, Room 4N222, Bethesda, MD 20892, USA. Email: desimonr@intra.nimh.nih.gov

Wholes to parts

Nick Donnelly
Centre for Visual Cognition, School of Psychology, University of Southampton, UK

One of the central issues underlying the psychology of visual processing is the question of whether it operates from wholes to parts or from parts to wholes. The wealth of neurophysiological data gained over the past 40 years showing early detection of object properties such as orientation and colour has led to the assumption that the psychology of visual processing follows a part-to-whole path. There are, however, findings that are difficult to explain within a framework of local to global processing. One significant illustration of whole over part processing is the perception of human faces by fellow humans. Years of experience through childhood and adolescence result in adults treating facial features as belonging to single objects and the formation of representations where wholes dominate over parts (Davidoff & Donnelly, 1990; Tanaka & Farah, 1993). This research has demonstrated quite clearly that holistic representations of faces are not mediated by prior analysis of parts (e.g., Suzuki & Cavanagh, 2003).

Some researchers have argued that faces are a special case, enabling them to be processed as wholes. In support of this argument reference is made to the particularly strong orientation specificity of face processing and the neural structures purported to be dedicated to face processing (e.g., the fusiform gyrus). The argument for the specialty of face processing is not, however, watertight. The evidence seems to point to expertise as being a significant factor in determining both inversion effects (Diamond & Carey, 1986) and activation of the fusiform face area (Gauthier, Skudlarski, Gore, & Anderson, 2000). It is possible to argue that only highly familiar, complex objects are processed as wholes, unmediated by an analysis by parts.

Chen's paper and those of others (e.g., Pomerantz, Sager, & Stoever, 1977) serve as a reminder that there is a wealth of evidence demonstrating that wholes are an important, perhaps crucial, unit of analysis when considering the psychology of visual processing more generally. A further example that illustrates this point is impossible objects, as these are often used as a test-bed for theories of visual representation (e.g., Donnelly, Found, & Muller, 1999; Hillstrom, Dror, Donnelly, & Rendell, 2004). It is often the case that impossible objects generate an initial feeling of plausibility followed by the sense that they are not possible; a good example of this is the Devil's tuning fork. Errors of believing impossible objects to be possible are usually attributed to faulty line-labelling or incomplete consistency checking between line-labels, where both errors indicate manifestations of local-to-global encoding schemes. Chen's paper suggests that, at least in some impossible objects, a different interpretation is possible. This can be demonstrated with reference to the Devil's torpedo tubes (Cowie & Perrott, 1993; see Figure 1). Here, participants have to judge whether two dots were on the same or different objects following a brief presentation. Most participants report that the dots are on the same object even though they are on objects that are physically disconnected. This illusion can persist over extended viewing, indicating that is unlikely that errors in line-labelling or incomplete consistency checking can really account for incorrect reporting. In addition, the illusion is difficult to understand if connectedness is the basic entry point for unit formation in visual perception. If connectedness and local analyses do not explain the phenomenology of the Devil's torpedo tubes, then Chen's theory might provide a more natural explanation of the illusion. The illusion is dependant on the proximity and parallelism of the neighbouring edges of the two spatially separate parts. Proximity and parallelism are both properties that Chen suggests

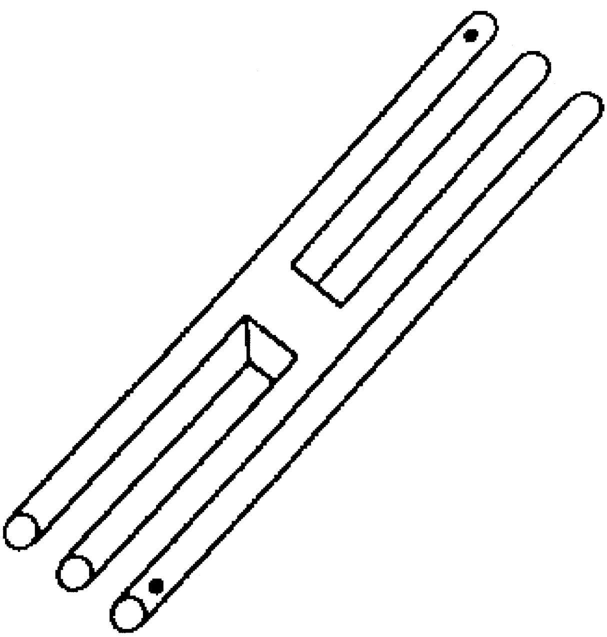

Figure 1. The Devil's tuning fork (from Cowie & Perrott, 1993, reprinted with permission).

are established early in processing. Consideration of the psychology of visual processing of a range of different objects does, therefore, suggest that wholes influence processing across different types of tasks and stimuli.

The novelty in Chen's analysis is reflected in the specifics of the proposed global to local scheme. For Chen, moving from global to local does not mean changing spatial scale (Navon, 1977) or finding geometric transformations to describe stimulus regularities (Dodwell, 1983). Rather it is in extracting invariant properties that can be described by different geometries. The first properties extracted are topological, where relationships between discrete objects are considered within a tolerance space. The important point is that topological properties are relatively stable over rubber sheet distortions and as such change our usual conception of ''shape''. The last properties to be extracted are those based on Euclidean space as these are the least stable of any geometry. While I take the mathematics underlying Chen's theory on trust, it seems that the experiments reported largely support the framework set out by this theory.

Whatever the utility of Chen's theory its impact will be greater if it is specified in more detail. Two related issues that seem to be of particular importance are noted here. First, theories describing how perceptual representations emerge over time assume either a blind acquisition of features or contingencies whereby representations of later features are guided by the influence of those extracted early in processing (e.g., Sanocki, 1993). Chen's theory is one of feature accumulation (where features are defined by different geometries) but it might also be constructed with time-course contingencies built into it. If so, what might these contingencies be and how might they influence processing? A second related issue is how perceptual groups defined by, for example, topology become ''anchored'' in Euclidean space. Chen envisages the experience of visual representation as a hybrid of all geometric descriptions grounded in some pseudo-Euclidean space (Dodwell, 1983). Chen does not, however, explain the implications of establishing coherent perceptual groups based on visual properties that can only be made explicit later in processing.

Defining the missing elements of Chen's theory would allow it to be treated in the same way as Marr's (1982) theory, which was used to reanalyse the breakdown of visual processing following brain damage (e.g., Humphreys & Riddoch, 1987). Even without further specification Chen's theory makes some testable neuropsychological predictions. For example, current research assumes that basic shape processing is intact when a patient passes the Efron test (Efron, 1968). In the Efron test, solid black rectangles of equal flux, where height and width can vary, are presented for patients to perform a same–different task. Chen's theory, however, suggests different levels of ''shape'' processing, like that of topological shape. An appropriate test of this theory would be showing patients topologically similar and dissimilar forms, as Chen has done in his tachistoscopic shape discrimination studies on typical adults. According to Chen's analysis there may be brain-damaged patients who are unable to perform the Efron test, but who can discriminate topologically similar from dissimilar forms. One candidate for this type of question would be DF, the visual form agnostic patient reported by (Milner & Goodale, 1995).

In fact Glyn Humphreys, Jane Riddoch, and I made some tentative steps to investigate this issue some years ago in the case of GK, a patient with separate bilateral lesions to the parietal cortex. Across a number of different tasks (two alternative forced choice and visual search) we contrasted his performance on target and distractor pairings matched on flux, but varying on shape (defined in Euclidean space) and/or topology. Although the data were never reliable enough to publish in an archive journal, there was some indication that he found discriminating rings from solid circles somewhat easier than squares from triangles. (For an account of some related data on the importance of closure to GK see Humphreys, Riddoch, Donnelly, Freeman, Boucart, & Muller, 1994.) Notwithstanding the case of GK, it may be a profitable exercise to test brain-damaged patients with marked problems in low-level visual processing on their ability to extract invariant properties consistent with different geometries.

In summary, Chen's paper does provide a case both for the psychology of visual processing being considered as going from global-to-local and that this transition should be made through the extraction of invariants features made explicit by different geometries. I find the first part of this statement easier to support that the second part. There is an abundance of evidence that wholes play an important role early in visual representation. Furthermore I have suggested that, if the correct conditions are in place, it may also be true not only in perceptual organisation but also in face processing and object representation as well. Whether Chen's specific scheme can be upheld is more problematic. Like most experiments, some of those performed and reported by Chen and his co-workers are open to other interpretations. Further refinement is required to make explicit the implications of encoding geometrical invariants early in processing for information that can only be made explicit later in processing. While there is no definitive evidence to support Chen's position, the suggestion that all relevant experiments can be understood within a single framework does need to be taken seriously.

REFERENCES

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Cowie, R., & Perrott, R. (1993). From line drawings to impressions of 3-D objects: Developing a model to account for the shapes that people see. *Image and Vision Computing*, *11*, 342–352.

Davidoff, J. B., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes. *Acta Psychologia*, *73*, 225–243.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107–117.

Dodwell, P. C. (1983). The lie transformation group model of visual perception. *Perception and Psychophysics*, *34*, 1–16.

Donnelly, N., Found, A., & Muller, H. (1999). Searching for impossible objects: Processing form and attributes in early vision. *Perception and Psychophysics*, *61*, 675–690.

Efron, R. (1968). What is perception? *Boston Studies in Philosophy of Science*, *4*, 137–173.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*, 197.

Hillstrom, A., Dror, I., Donnelly, N., & Rendell, N. (2004). Impossible objects: A testbed for understanding object representation and processing. *Manuscript in preparation*.

Humphreys, G. W., & Riddoch, M. J. (1987). The fractionation of visual agnosia. In *Visual object processing: A cognitive neuropsychological approach* (pp. 281–306). Hove, UK: Lawrence Erlbaum Associates Ltd.

Humphreys, G. W., Riddoch, M. J., Donnelly, N., Freeman, T., Boucart, M., & Muller, H. (1994). Intermediate visual processing and visual agnosia. In M. J. Farah & G. Ratcliff (Eds.), *The neuropsychology of high-level vision* (pp. 63–101). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford, UK: Oxford University Press.

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

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

Sanocki, T. (1993). Time course of object identification: Evidence for a global-to-local contingency. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 878–898.

Suzuki, S., & Cavanagh, P. (2003). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 901–913.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A*, 225–245.

Please address all correspondence to: N. Donnelly, Centre for Visual Cognition, School of Psychology, Southampton SO17 1BJ, UK. Email: n.donnelly@soton.ac.uk

Wholes and parts: A chicken and egg story?

Anthony Hayes

Department of Psychology, The University of Hong Kong, China

Many who investigate human vision implicitly subscribe to a theory of vision whereby the visual system engages in decomposition followed by reassembly. One version of this view is that the visual system takes the retinal image and breaks it down into many fragmented image attributes, such as line segments and corners, and motion and texture and colour, all of which are independently processed. At some late and ill-defined stage, the attributes are brought together. What decomposition and independent processing, followed by reconstitution, achieves, or can achieve, is usually not satisfactorily addressed by those who hold this view. Furthermore, data are published from time to time that are not easily reconciled with this view (e.g., recent work of myself and colleagues; Hayes, 2000; Ross, Badcock, & Hayes, 2000; see also Burr, 1999, 2000).

David Marr (1982) possibly articulated the most sophisticated answer to the question ''Why fragment the visual image merely to expend large amounts of neural resources in piecing the fragments together again?'' when he reminded us that it is local structural variations in the image that allow us to isolate objects in images—starting, in his view, most notably with edges (Marr & Hildreth, 1980)—but that these structural variations are *implicit* in the grey-scale (and colour) array. A necessary early goal of vision is to make the structural variations explicit. For Marr, fragmentation was a consequence of explication, and binding of the fragments was thought in part to be achieved by processes that take place in scale-space where adjacencies over a range of scales are exploited to automatically associate, perhaps via an heuristic rule, the elements.

Marr's account has never been an especially satisfactory one, if for no other reason than that for nearly all structural features of a natural image, there are changes in position and orientation at any given location as one transits scale (see, e.g., Field, Hayes, & Hess, 1993). These changes in position and orientation are often substantial, and suggest that to achieve appropriate association of image fragments a scale grouping rule is likely to be insufficient. One may need, Gestalt like, the whole as a guide to sum the parts. Lin Chen has the remarkable view that this may be precisely what vision does.

Chen casts the above view that visual processing acts from local to global analyses—such as Marr's view of early vision—and the view that the visual system proceeds from global to local, as a Great Divide. Chen's view, that wholes are coded prior to analysis of their separable properties or parts, is a view inherent in the Gestalt conception of perceptual organization. Specifically, Chen's view is supported by evidence that has resulted in a theory of *early* topological perception in which a primitive and general function of vision is the perception of topological properties. In this conception, a hierarchy of structural properties is proposed where primacy is given to topological perception, based on physical connectivity.

This view that the visual system is primarily concerned with finding what is connected to what, and not with attribute decomposition, is appealing from a consideration of a big-picture analysis of what vision is for, but more importantly, it is backed up by a range of recent experimental work, much of which comes from Chen and his colleagues' laboratories (e.g., Chen, Zhang, & Srinivasan, 2003; Zhuo et al., 2003).

While on first inspection Chen's view seems to be back-to-front, it accords with modern anatomical findings about the visual system: There is great emphasis in the visual pathway on feedback. The function of feedback is often thought of as being to moderate processes earlier in the pathway, but there is no good account of what the shape of this moderation may be. In Chen's view of global-to-local perceptual processing, the topological structure of the whole acts as a primary moderator. Indeed, the topological structure—the whole—may be a necessary condition to allow the parts to sum. The topological approach, while at one level incompatible with what may be regarded as the traditional computational approach, can nevertheless be thought of as a computational theory (in the sense of Marr) in which the overriding constraint is invariance in topological transformation. As with Marr's use of the idea of a constraint, the topological approach emphasizes the early impact of topological invariants on vision. In this conception the whole is indeed ''greater than the sum of the parts'', but also, the whole has precedence over the parts.

REFERENCES

Burr, D. C. (1999). Vision: Modular analysis—or not? *Current Biology*, *9*, R90–R92.

Burr, D. C. (2000). Motion vision: Are ''speed lines'' used in human visual motion? *Current Biology*, *10*, R440–R443.

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Chen, L., Zhang, S. W., & Srinivasan, M. V. (2003). Global perception in small brains: Topological pattern recognition in honeybees. *Proceedings of the National Academy of Sciences, USA*, *100*, 6884–6889.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Good continuation and the association field: Evidence for local feature integration by the visual system. *Vision Research*, *33*, 173–193.

Hayes, A. (2000). Apparent position governs contour-element binding by the visual system. *Proceedings of the Royal Society, London, Series B*, *267*, 1341–1345.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society, London, Series B*, *207*, 187–217.

Ross, J., Badcock, D. R., & Hayes, A. (2000). Coherent global motion in the absence of coherent velocity signals, *Current Biology*, *10*, 679–682.

Zhuo, Y, Zhou, T. G., Rao, H. Y., Wang, J. J., Meng, M., Chen, M., Zhou, C., & Chen, L. (2003). Contributions of the visual ventral pathway to long-range apparent motion, *Science*, *299*, 417–420.

Please address all correspondence to: Anthony Hayes, Department of Psychology, The University of Hong Kong, Pokfulam Road, Hong Kong, China. Email: ahayes@hku.hk

Temporal tolerance circumscribed

Mark A. Elliott
Temporal Factors and Psychophysics Laboratory, Department Psychologie, Abteilung für allgemeine und experimentelle Psychologie, Ludwig-Maximilians Universität, Munich, Germany

Chen's topological theory of perceptual organization includes the idea that equivalences in spatial organization and the organization of subjective experiences of time may be derived by means of common expression in terms of topological tolerance spaces. A critique is offered of some foundational concepts related to Chen's suggested organization of subjective time, culminating in the proposal that the organization of time may be best considered with reference to some medium or organization, in this instance the dynamic patterning of nervous system activity. Direct equivalences are then identified between dynamic nervous system activity correlated with spatial and temporal organization. These nevertheless raise the question of whether we need to consider organization of subjective temporal experiences in order to evaluate the idea that variations in process timing is of itself a common mode of expression for spatial and temporal organization.

Event timing and the relative latency of an associated perceptual response is never far from the idea that perceptual priority is allocated to stimuli according to the extent of topological invariance across transformations: It is of course central to Chen's topological approach to perceptual organization in the sense that it is paradigmatic; one cannot infer variations in processing precedence according to topological variance/invariance without manipulations in the timing of stimulus events. But is ''psychological time'' in a more general sense of any theoretical relevance to Chen's theory of perceptual organization in space?

Chen points out that variability in subjective time relative to the actual timing of a physical event results in the formation of discrete temporal ranges over which the event is perceived or estimated to occur. Range formation applies to the experience of (apparent) simultaneity, temporal order, and/or the *psychological present*: Chen argues that equivalences between the organization of subjective time and the organization of perceptual space may be drawn by common reference to topologically definable tolerance spaces. A tolerance space is a description of the range over which detailed variation in a given stimulus is ignored. The degree of tolerance in space is in principle describable by some function that describes a set of items according to some spatially extended (but nonetheless definable) organization in space. In a similar fashion, a ''temporal tolerance space'' might be assumed to be some function of the range of variability in estimated time relative to the actual timing of an event. What's more, equivalence between spatial and temporal tolerance spaces implies psychological time to be possessed of *Prägnanz* in the same basic sense as certain classes of organization in space may possess it.

Before general consideration of the notion and implications of equivalence between temporal and spatial organization it seems necessary to take a particular ''position'' regarding analogies between spatial and temporal organization and thereafter to identify and resolve two problematic aspects of Chen's suggested organization in time. With respects to taking a position on the possibility for space–time analogies it should be noted that the general analogy is suggestive at first sight, but time is different from space. It is one-dimensional (i.e., it is topologically poor), it is oriented in the time-flow direction and can obviously not easily serve for simultaneous representation of series of nonsimultaneous events. Besides this (and as is repeated later as a major theme of this commentary), the indications are that relations between physical time and phenomenological time are far from consistent. In fact they may be inconsistent to such an extent that it may be difficult to do more than intuit a generalizable structure in phenomenological time and then perhaps better to consider those nontemporal factors to which any pattern of inconsistency ultimately refers. Consequently, it will be argued that any position on the structural relations between (the representation of) space and time must necessarily adopt a flexible outlook, and it is only from a position of flexibility that there exists any substantive basis for further consideration of the notion of (or put another way what actually is) ''temporal structure'' and how this structure might enter into the topological calculus.

Nevertheless by assuming temporal structure we then encounter our first problems of definition, and these problems relate specifically to the validity of the idea, advanced by Chen (and formerly by Pöppel, 1985, 1997) that events in time are related hierarchically. The concept of a hierarchy entails reference to a system of elements arranged (i.e., related) into graded (i.e., ranked) order. Moreover, it presupposes grading to be defined by some relational, perhaps inherent measure. [As an aside, I think it is probably true to say that the hierarchical representation of relations between tolerance spaces (and what Chen has in mind by reference to the idea of a hierarchy) may be best conceptualized in terms of a ''dominance hierarchy'' in which, for example, the relations between three tolerance spaces (X, ξ_1); (X, ξ_2); (X, ξ_3) may be best represented as $(\xi_1 < \xi_2) < \xi_3$. This may not be quite the same idea as put forward by Pöppel, who certainly implies that all temporal experiences are in some way graded relative to one another, but does not seem to define the precise way in which temporal experiences come to constitute their hierarchy.]

However, defining items in terms of some continuum simply because they describe a range of variation along that continuum does not guarantee that the items form a set. If selecting items for set inclusion solely on the basis of their apparent position along such a continuum the resulting set may lack criterion-related validity. The set may also lack internal validity, particularly when the items concerned consist of a number of attributes and when these attributes are ignored for the purposes of set inclusion. Problematic validity may cause false gradings to arise. An example of the fallacy of grading or ranking simply based upon some measure of magnitude concerns the U-shaped sensitivity functions derived from metacontrast masking. The existence of nonlinearities of this character would necessitate consideration of the two disjunct periods of good target discrimination performance as of separate orders simply because one period occurs ''before'' and is of a lesser magnitude, whereas one occurs ''after'' a period of poor performance and is of a higher magnitude. While it may very well be the case that the underlying processes differ, the problem is one of definition; one might not logically be able to define the period of poor performance as a special case if the performance series is considered in terms of three qualitatively different ranges. The implications are subtle but impact at both a systems and at a metatheoretic level of analysis. In the first instance performance might be considered better in terms of three successive but nonetheless related systems as opposed to a single system with nonlinear characteristic. Consequently, at the metatheoretic level the notion of perception as essentially a linear, feedforward system may be supported simply because nonlinearities can be argued away. It may be precisely this problem that Pöppel (1997) seeks to avoid by introduction of the notion of hierarchical gradings of temporal perception. My point, nonetheless, is that it may be misleading to consider all forms of operation in psychological time in this frame of reference.

The notion of a hierarchy may also be somewhat misleading for another reason: The timing of psychological events such as the patterning of estimates in duration discrimination can be shown to obey a quantal staircase law of successive doublings, appearing as a concomitant of ranges whose expansions relate to physical time as a function of Weber's law (see Kristofferson, 1990). Consequently, one can say that at certain intervals, subjective durations will differ in magnitude but these differences in magnitude, while manifest in time, do not refer intrinsically to differences in some magnitude of elapsed time, but of something else that comes to determine their nonlinear characteristic. It might be the case that in all areas except the study of time itself, relations expressed in terms of some temporal characteristic may experience a fundamental problem of reference with respects to expression in terms of ordinal or indeed nominal units of physical time. A more cautious approach to claims of the existence of a hierarchy of psychological time would be to avoid too much emphasis upon arbitrary subdivision in time and consider ''to what it is'' we are referring by reference to variations in timing of particular, temporally defined experiences or events.[1]

[1] At the very least, if we are to question the applicability of the concept of a ''temporal hierarchy'', we might do so on the basis of its inherent constructivist implications, which tend to oppose the global perspective taken by Chen with regards to perceptual organization in space.

A similar problem accompanies a second aspect of Chen's position on the possibility for spatiotemporal equivalence, notably a common level of description relating various types of subjective time. Evidence certainly exists to suggest a common level of description for different perceptions of stimuli in time, for example the apparent concomitance of the perception of apparent motion with that of precise temporal order judgement, which breaks down as stimuli, perceived as a single moving point under different temporal conditions come to be judged as two successive points (e.g., Westheimer & McKee, 1977). However, and as alluded during earlier discussion of the notion of a temporal hierarchy, it might be difficult to properly quantify links between the experiential characteristics of temporally defined perceptions (such as the perception of (stimulus) simultaneity and the experience of phenomenal ''presence''). In this respect, particular caution must be exercised with respects to subjective time for two reasons: Firsty, a consistent and comparable mapping of subjective to objective time cannot be assumed across all classes of temporal experience and secondly, physical time is not a stimulus in the conventional sense of the word. Consequently, and given well-known, but as yet unexplained variations in subjective relative to objective time, it may be difficult to rely upon physical time as a standard.

As a consequence it may be more useful to remain reliant upon the empirical independent definition of each class of temporal experience and consequently in terms of discrete rather than interrelated ranges of variation. In this way, it seems reasonable to consider different types of temporally defined experience in a similar fashion to different types of organization in space, that is to say, related on a metatheoretic level, perhaps in terms of their Gestalt or Prägnant characteristics while at the same time subject to lawful, stimulus related variation. What's more, as my colleague Hans Geissler has observed, ''establishing measures of subjective time involves knowledge of the *total perceptual-cognitive structures* concerned (quite in analogy to subjective spatial relations)'' (H. Geissler, personal communication, emphasis added). This point also applies incidentally to the definition of tolerances, which can only be adequately defined in relation to the percept as a whole. Given that we can currently only speculate concerning the range of processing variation that brings about subjective experiences of time it seems once again sensible to conclude that relating experiences of psychological time may be premature and may suffer from the same lack of criterion-related validity as the idea that types of experience in time should be organized hierarchically.

The establishment of external validity must come about as a continued function of the sensitive covariation of experience and behaviour with stimulus-related variation, as is the substance of time estimation research and what is coming to be referred to as ''process psychophysics''. Nevertheless, and as counselled above with respects to the definition of a hierarchy of temporal experience, a common definition for events or experiences nominally or ordinally related in time might be best addressed by reference to something outside of the temporal domain. Henri Bergson was amongst the first to argue on logical grounds that we should consider our conceptualization of ''duration'' (i.e., extension in time) to draw reference to our conceptualization of ''extension in space''. Put in synopsis, Bergson's position was that '' 'dureé pure' or 'pure duration' is wholly qualitative. It cannot be measured unless symbolically represented in space'' (Bergson, 1910, p. 104).

However, even by the turn of the twentieth century Bergson's position relative to the ontological status of temporal experience was not new. In fact there existed in the theoretical reflections of both Gustav Theodor Fechner (1860) and Karl Ernst von Baer (1864) an empirically operationalizable proposal concerning the psychophysical representation (really the physical representation as psychophysics refers to the distribution of energy in a spatially defined system of references) of subjective duration based upon a near identical conceptualization to that given by Bergson. For von Baer ''time'' was made up of a series of subjective moments, which are equivalent to the briefest time units in which the world appears stationary. In the theme of von Baer and some 70 years later, Brecher (1932) discriminated human moment to be in the order of $\frac{1}{18}$ of a second because pictures, sounds or taps to the skin can only be discriminated if applied at a tempo slower than this rate. In contrast to the human moment, Brecher observed that four tactile stimulations per second with a rod on the belly of a snail compels it to attempt to crawl upon a nonexistent coherent surface which it perceives as a simultaneity (in spite of the considerable delays between application of the stimulating rod). This indicates that the receptor time of the snail has a tempo of three to four moments per second. Consequently, events in the snail's world will appear much faster than our own experience of them although it might be conjectured that its own motions would seem no slower to the snail than ours do to us. On these bases, it was reasoned the experience of events in time to be both relative and fundamentally determined by structural factors such as the overall length of life and the complexity of the nervous system. In other words subjective time was considered to be a biological property, in part instantiated in the nervous system.

The implications of this position are twofold: On the one hand the desired common level of description upon which substantive parallels might be drawn between temporal and spatial coding might be lie in common operating characteristics within the nervous system. The visual system, upon which basis the great majority of our understanding of spatial coding is derived is very well researched and methods for deriving measures of visual performance, both physiological and psychophysical, offer a level of precision upon which mathematical models of perceptual organization might be designed and tested. Similarly, the question of whether different aspects of temporal experience prove to be governed by the same laws of process operation might be formulated as a set of empirical questions targeted towards a precise specification of what

Pöppel (1997, and by association of ideas, Chen) refers to in terms of the ‘‘neurocognitive mechanisms’’ or the common neural algorithms subserving temporal perception.

What of these mechanisms and how to develop substantives links between nervous system activity and psychophysical performance? Fechner (1860) proposed ‘‘inner psychophysics’’ to comprise cascades of neuronal oscillations, which might serve either singularly or in cascade to ensure subjective time proceeds in accord with physical time. Echoing this proposal, Pöppel (1997) has suggested temporal perception as reducible to patterns of oscillatory activity, going further to suggest the appearance of oscillatory activity at around 33 Hz as of particular significance. What seems promising for an evaluation of equivalences between spatial and temporal organizations is evidence that presentation of spatial organizations of the sort described by Gestalt psychologists correlate with the synchronization of interneuronal firing rates at firing frequencies in a similar range to those described by Pöppel (i.e., in the 30–60 Hz range; see Gray, 1999 for a discussion on this issue). Nevertheless, reduction of the specification of spatial and temporal organization in terms of process timing raises a number of issues. Not the least of these is a proper understanding of the extent to which we should consider the dynamics of nervous system activity to relate better to the idea of organization as it is understood to imply the ‘‘act of organizing’’, or, as it is intended to refer to, the ‘‘state of organization’’. Intuitively, the former definition suits the dynamic character inherent in the idea that ‘‘organizations’’ come about as a function of oscillatory activity in the brain. However, and of importance for the aforementioned consideration, is the extent to which the tendency for oscillations to occur at one or other frequency may be best explained in terms of synergism. In other words, to what extent a given neuronal oscillation should be best considered as some emergent property or indeed some reciprocal of a multiplicity of static and inbuilt physical constraints such as neural conduction latencies, refractory periods the extent of the neuronal substrate engaged in a given information processing task, and so on, and indeed, not forgetting determination as a function of other patterns of oscillatory activity.

Issues such as these require resolution before it can be properly decided that temporal and spatial organizations can be considered comparable at the level of nervous system activity. What then of the earlier posed question of the theoretical relevance of psychological time to Chen’s theory of perceptual organization in space? Consistent with the ideas of Pöppel to whom Chen refers, our journey from psychological time has lead to consideration of psychological time as most usefully expressed as patterns in process timing. In this respect, our reduction of subjective time bears striking similarity to a similar reduction of spatial organization. Both can be described in terms of process states that express themselves in oscillatory activity at particular (and closely matching) frequencies. On this basis, and irrespective to the commonality of the neuronal algorithm concerned, we are forced to ask whether or not variations in spatial organization evaluated relative to covariations (and critically the variability) in associated patterns of process timing would not be, of itself, a sufficient measure of the relations between spatial and temporal organization? Perhaps it is upon this basis that tolerance spaces might be directly computable and quantitative equivalences between temporal and spatial organizations revealed.

ACKNOWLEDGEMENTS

I am indebted to Hans Geissler for numerous discussions on these and related issues and to Viktor Sarris for clarification of Brecher's contribution to the study of the perceptual moment. Grateful thanks are expressed to Stanislava Antonijević for proofreading this text. The author's work is supported by deutsche Forschungsgemeinschaft project grants EL 248/1 and El 248/2.

REFERENCES

Bergson, H. (1910). *Time and free will: An essay on the immediate data of consciousness* (F. L. Pogson, Trans.). London: George Allen & Unwin.

Brecher, G. A. (1932). Die Entstehung und biologische Bedeutung der subjectktiven Zeiteinheit – des Momentes. *Zeitschrift für vergleichende Physiologie*, *18*, 204–243.

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Fechner, G. T. (1860). *Elemente der Psychophysik*, *Vol. 2*. Leipzig, Germany: Breitkopf & Haertel.

Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, *24*(7–9), 31–47.

Kristofferson, A. B. (1990). Timing mechanisms and the threshold for duration. In H.-G. Geissler, in collaboration with M. H. Müller & W. Prinz (Eds.), *Psychophysical explorations of mental structures: Selected papers of the centennial symposium in honor of Gustav Theodor Fechner* (pp 269–277). Toronto, Canada: Hogrefe & Huber.

Pöppel, E. (1985). *Mindworks: Time and conscious experience*. Orlando, FL: Harcourt, Brace Jovanovich.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, *1*(2), 56–61.

Von Baer, K. E. (1864). Welche Auffassung der lebenden Natur ist die richtige? Und wie ist diese Auffassung auf die Entomologie anzuwenden? In K. E. von Baer (Ed.), *Reden, gehalten in wissenschaftlichen Versammlungen und kleinere Aufsätze vermischten Inhalt* [the first part given as an oral lecture in 1860] (pp. 237–284). St. Petersburg, Russia: H. Schmitzdorf.

Westheimer, G., & McKee, S. P. (1977). Perception of temporal order in adjacent visual stimuli. *Vision Research*, *17*, 887–892.

Please address all correspondence to: Mark A. Elliott, Department Psychologie, Abteilung für allgemeine und experimentelle Psychologie, Ludwig-Maximilians Universität, München, Leopoldstr. 13, D-80802 Munich/München, Germany. Email: elliott@psy.uni-muenchen.de.

From topological perception to distributed cognition

Jiajie Zhang

School of Health Information Sciences, University of Texas at Houston, USA

Chen's theory of topological perception not only presents a major challenge to many currently adopted views of perception but also has profound implications for the research on higher level cognition. In this commentary I will discuss the impact of the major ideas of topological perception on a new emerging paradigm in cognitive science. This impact can be briefly summarized in the following argument: Chen's theory of topological perception provides a unique and systematical mathematical and theoretical treatment for Gibson's theory of ecological perception, which provides key theoretical concepts and principles for the studies of external representations, which are in turn central to the theory of distributed cognition—the new emerging paradigm in cognitive science. In the following paragraphs, I will elaborate on this multilevel relationship.

Traditional cognitive science has focused on the structures and processes of mental representations, which are internal representations in people's heads. The achievements along this direction have been very impressive. However, this traditional approach has run into some difficulties in the study of people's behaviour in complex social, cultural, and physical environments. Most studies in traditional cognitive science either completely ignored external representations or, when taking external representations into account, often failed to separate them from internal ones. Thus, these studies often mistakenly equate external representations to internal representations. As noted by Kirlik, Plamondon, Lytton, and Jagacinski (1993) and Suchman (1987), this confusion often leads one to postulate unnecessary complex internal mechanisms to explain the complex structure of the wrongly identified inernal representation, much of which is merely a reflection of the structure of the external representation (Simon, 1981).

To address this confusion, several lines of research in cognitive science have emphasized the structures of the environment and people's interacions with them. The distributed cognition approach explores how cognitive activity is distributed across internal human minds, external cognitive artifacts, and groups of people, and across space and time (e.g., Hutchins, 1995a, 1995b; Norman, 1993; Zhang, 1997; Zhang & Norman, 1994). According to this approach, people's intelligent behaviours result from interactions with external representations and with other people. The representation of a distributed cognitive task is distributed as a system of distributed representations with internal and external representations as two indispensable parts. External representations are neither mere inputs and stimuli to nor mere memory aids to the internal mind. Rather, they are intrinsic components of many cognitive tasks and they guide, constrain, and to some extent determine cognitive behaviour. Another related approach, situation cognition, makes similar arguments (e.g., Barwise & Perry, 1983; Clancey, 1993; Greeno & Moore, 1993; Lave, 1988; Suchman, 1987).

External representations, according to Zhang (1997; Zhang & Norman, 1994), are the knowledge and structure in the environment, as physical symbols, objects, or dimensions (e.g., written symbols, beads of abacuses, dimensions of a graph, etc.), and as external rules, constraints, or relations embedded in physical configurations (e.g., spatial relations of written digits, visual and spatial layouts of diagrams, physical constraints in abacuses, etc.). As a major component of the theory of distributed cognition, external representations are theoretically motivated by and built upon Gibson's (1966, 1979) theory of ecological psychology. According to Gibson, the environment is highly structured with invariant information in the spatial and temporal patterns of optic arrays, and the invariant information in the environment can be directly picked up by perceptual systems without the mediation of memory, inference, deliberation, or other mental processes. The information in the environment is sufficient for perception and action because it is sufficient to specify all objects and events in the environment. In addition, the end product of perception is not an internal representation of the environment but the invariant directly picked up from the environment. Affordances are examples of the invariants that can be directly perceived and used without being interpreted and formulated explicitly.

Gibson's ecological psychology, despite its complexity, does not have a solid mathematic foundation, which is required for any further advancement of the theory. Chen's theory of topological perception just provides such a needed foundation. Specifically, the structural and functional hierarchy of perception and the notion of tolerance space for perception, among other things, provide a nice mathematic treatment for invariant information, which is the central piece of Gibson's ecology psychology (for details, see Part V of the target paper). In the theory of topological perception, the invariant structures in the environment are the invariant properties under various geometrical transformations such as symmetrical, affine, projective, and topological transformations. The one-to-one mapping requirement, which is a difficulty for linking the one-to-one mapping in geometry to the sometimes not one-to-one mapping in ecological array, is handled by applying the global characteristic of tolerance spaces that essentially removes the local assumption of ''one-to-one'' mapping. Chen's theory not only provides a structural description of invariant information but also gives a process explanation of direct perception. Under this theory, direct perception is the perception of global invariants without the mediating processes of local feature analysis. This is because the perception of global features occurs before the perception of local features and because the measure of local features depends on the global properties of objects.

Chen's theory of topological perception is an important contribution to not just the research on perception but also the research on higher cognition as discussed above. More importantly, the notion of directly perceivable information in the environment is not an ad hoc explanation of the phenomena in distributed cognition; rather, it has been the driving force in the emergence of a new paradigm called distributed cognition. Chen's theory has opened new lines of research that are worth of attention from not just researchers in perception but also researchers in cognitive science in general.

ACKNOWLEDGEMENTS

Preparation of this commentary was supported by a grant from the Human-Centered Computing and Intelligent Systems Program, National Aeronautics and Space Administration, Grant No. NCC 2-1234.

REFERENCES

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: MIT Press.

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Clancey, W. J. (1993). Situated action: A neuropsychological interpretation response to Vera and Simon. *Cognitive Science*, *17*(1), 87–116.

Gibson, J. J. (1966). *The senses considred as perceptual systems*. New York: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. New York: Houghton Mifflin.

Greeno, J. G., & Moore, M. J. (1993). Situativity and symbols (Response to Vera and Simon). *Cognitive Science*, *17*, 49–59.

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge, MA: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speed. *Cognitive Science*, *19*, 265–288.

Kirlik, A., Plamondon, B. D., Lytton, L., & Jagacinski, R. J. (1993). Supervisory control in a dynamic and uncertain environment: A process model of skilled human–environment interaction. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*(4), 929–952.

Lave, J. H. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.

Norman, D. A. (1993). *Things that make us smart*. Reading, MA: Addision-Wesley.

Simon, H. A. (1981). *The sciences of the artificial* (2nd ed.). Cambridge, MA: MIT Press.

Suchman, L. A. (1987). *Plans and situated action: The problem of human–machine communication*. New York: Cambridge University Press.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, *21*, 179–217.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, *18*, 87–122.

Please address all correspondence to: Jiajie Zhang, School of Health Information Sciences, University of Texas at Houston, 7000 Fannin, Suite 600, Houston, TX 77030, USA. Email: Jiajie.Zhang@uth.tmc.edu

Complementarity as a generative principle in visual perception

Ernst Pöppel

Institute of Medical Psychology, Ludwig-Maximilians-Universität, Munich, Germany

A theory of visual perception has to explain a vast number of observations. Let me name just a few, and by doing so, I want to propose a direction for further considerations which in my view would necessarily have to be based on experimental observations and theoretical concepts presented by Lin Chen. He starts his paper with a Chinese proverb, i.e., that everything is difficult in the beginning. Let me complement this with a German proverb: Who does not know the goal cannot find the way.

What are such goals? One has to address (1) the maintenance of identity of a visual percept throughout time (at least for some time). How is it possible that a percept can stay the same in spite of the rapidly changing activities of local neuronal elements that represent the neuronal basis of the percept? This question leads to the next one, (2) how a percept is on the one hand maintained over time, and how on the other hand this percept is replaced after some time by another percept. What are the neuronal mechanisms that both create perceptual stability and perceptual dynamics? One has to deal (3) with constancy mechanisms like size constancy, colour constancy, or constancy of brightness throughout the visual field. Do constancy mechanisms contribute to the phenomenal identity of a visual percept? One has to discuss (4) the complementarity of local features and topological invariants both being necessary (as I see it) to generate a visual percept. One has to look (5) for mechanisms of complexity reduction in the temporal domain, i.e., the creation of order presumably already on the precortical level. In what follows I present some ideas that are motivated by the work of Lin Chen and that give a new meaning to some of my own previous work, and these considerations lead me to suggest complementarity as a generative principle in visual perception. I start with some observations.

If we perceive an object with foveal and depending on the size of the object with more or less perifoveal exposition, the first step of processing will be transduction of local stimuli on the level of the photoreceptors. As the distribution of photoreceptors on the retina is extremely inhomogeneous, and as an object is usually defined by regions of different brightness (objects are usually not characterized by monochromatic surfaces of identical flux), the transduction time will be different for different retinal regions that define the object. Brightness differences and inhomogeneity of the retina consequently result in temporal indeterminacy already on the level of the bipolar cells and consequently on the next processing levels like the ganglion cells or the geniculate neurones, i.e., local activities of neurones that carry the information to define the visual object show some unpredictable temporal variability. Under normal viewing conditions one has to expect a temporal uncertainty of at least 10 ms (Pöppel, Schill, & von Steinbüchel, 1990).

How could this problem be overcome, if it is at all a problem (what I am inclined to believe)? How could reduction of complexity work that is introduced by different transduction times and the retinal architecture? Before addressing this question let us look at another problem, i.e., at constancy of brightness throughout the visual field. Although the retina under photopic adaptation conditions is less sensitive towards the periphery, apparent brightness of any object is independent of the position of the visual field (Pöppel & Harvey, 1973). Under photopic adaptation conditions the operative range of the retina up to 60 degrees eccentricity is approximately 2.5 to 3 log. units. However, in the suprathreshold range apparent brightness corresponds directly to the physical intensity of the stimulus, independent of its position in the visual field. Thus, a homogeneous surface may be constructed already on the retinal level and this homogeneous surface that can be looked at as a sheet onto which at a further processing stage objects are ''projected''. Perceptual objects themselves are apparently free from the retinal inhomogeneity. Thus, there is complementarity between early stages of processing that create a homogeneous perceptual surface, and the neuronal machinery that is responsible for the genesis of the percept.

Let me come back to complexity reduction, which appears to be necessary because of the temporal problems due to biophysics and the architecture of afferent projections. What could be a solution for this problem? On the basis of some psychophysical and neurophysiological observations I would like to suggest that afferent information in the visual channel triggers oscillatory activities latest on the level of the lateral geniculate nucleus. Technically speaking these oscillatory activities correspond to a relaxation oscillation characterized by an immediate (or at least very fast) entrainability. One period of such a relaxation oscillation represents a hypothetical system state in the processing cascade. A system state is characterized by the fact that the before–after relationship is not defined; information within a system state is treated as cotemporal. With such cotemporal and thus atemporal zones the brain can create on a presemantic level frames or temporal processing windows within which temporally and spatially distributed activities can be integrated (Pöppel, 1997).

I would like to propose that system states of this kind could be used for complexity reduction. A neuronal sheet is created within which a functional link can be established between neighbouring elements. I want to stress again, that this occurs on a presemantic level. An early sketch of surfaces is provided by such stimulus-triggered activities, and thus overcoming the temporal uncertainty. (Obviously, some other aspects have to be considered here like the first-come first-win principle, giving rise to system states in the neuronal net at the geniculate level, which requires lateral interactions between the local elements.) Is there some evidence for this idea of an automatic complexity reduction? In fact there is quite a lot of evidence for neuronal processes being characterized by system states in the temporal domain of 30–40 ms like observations on single unit behaviour (Podvigin, Jokeit, Pöppel, Chizh, & Kiselyeva, 1992), or measurements on the initiation of eye movements (Pöppel & Logothetis, 1986). The response characteristics of visually triggered movements of such kind is interestingly not unimodal, which one would expect if visual information would be processed continuously, but typical response histograms are multimodal with temporal intervals between the modes of 30–40 ms. This phenomenon cannot be explained by a theory that is based on continuous processing of visual information; we are forced to conclude that an afferent stimulus elicits an excitatory process with periodic characteristics.

Reiterating again one point: System states are created automatically, in a presemantic way, i.e., independent of what it is that is processed in the visual pathway. System states simply provide a logistical frame for further processing. I think it is of importance to distinguish between those neuronal mechanisms that are responsible for the implementation of content—''what it is'' that I see—and those logistical or service functions like the temporal control function that provide a processing frame; we are confronted here again with the principle of complementarity.

The identification of complementary operative modes becomes also important if we try to link the hypothetical system states to the feature binding of a further processing stage. Binding of local features that establish for instance belongingness like surfaces or connectivities (i.e., edges) are conceived of being involved in the genesis of a percept (e.g., Engel, Fries, & Singer, 2001). Binding is a hypothetical mechanism responsible for ''what it is'' that is perceived. How could what I have hypothesized on system states be related to binding? I want to suggest that binding as it is discussed is a further processing step that operationally dwells on prior processing, i.e., on already established system states. There is already a general structuring or ordering from which a further mechanism might select. This further mechanism chooses from the temporally prestructured activities on the precortical level and constructs—necessarily by some top-down control—belongingness, which is expressed in the synchronization of distributed elements. However, an independent selection mechanism has to be active to choose those elements that belong together to give rise to a specific percept. Without such a semantic component, or an abstract representation of an object, it is difficult to imagine how binding could work. One has to address the question of who does the binding or the selection.

Let me turn now to another temporal process, which again is constituted presemantically, a process of temporal integration. It has been demonstrated by a number of psychophysical and neuropsychological experiments that the human brain provides a temporal platform of just a few seconds, and some neurophysiological observations support the existence of such a temporal platform with a duration of 2–3 s (Pöppel, 1997). I would like to repeat just a few examples. The reproduction of visual or auditory stimuli of different duration can be done veridically and with small variance up to approximately 2–3 s, but not beyond. Ambiguous visual figures switch automatically with spontaneous alteration rates of approximately 3 s. The shift of apparent patterns in binocular rivalry follows the same temporal dynamics representing each pattern just for a few seconds. Interestingly, automatic temporal segmentation in this time domain applies also to movement control, speech segmentation, or working memory.

This presemantically instantiated temporal window appears to be central for the representation of percepts. I would like to submit that a visual percept will be presented for just a few seconds onto this temporal stage. The most basic feature of an object is its phenomenal identity. What are potential mechanisms that allow the maintenance of perceptual identity? I think it is the main function of the constancy mechanisms to allow the construction of perceptual identity on a phenomenal level. I mentioned already constancy of brightness throughout the visual field. I should mention in particular also colour and size constancy. Let me have a closer look at size constancy. There are several mechanisms for size constancy, but I am looking at the one that is mediated by information coming from the extraocular muscles. It has been shown that information from the vergence and the accommodation system are used for this kind of size constancy, and it could be shown that this information results in a transformation of a central coordinate system (Pöppel, 1988). Thus, the mechanism of size constancy does not work on the perceptual object itself, but presemantically on an extrastriate module representing a topological map of the visual world around us, which can be modulated. Objects are represented—this is the hypothesis—within a topologically invariant coordinate system and apparent constancy of size as a result contributes to perceptual identity of an object. Let me stress again: The distinction between neuronal mechanisms on the presemantic level from those that allow the construction of content—of ''what it is'' that I see—and thus their complementary action is essential for this reasoning.

Let me just give one more example to exemplify the distinction between presemantic processing and content-oriented perception. The well-known phenomenon of residual vision or ''blindsight'' can also be interpreted with respect to a dissociation between the ''what it is'' that I see and basic neuronal logistics (Pöppel, Held, & Frost, 1973). The loss of neuronal processes in the occipital cortex destroys the possible content of visual perception; however, basic presemantic mechanisms are still in operation that allow spatial orientation or a subliminal detection. The lesion opens a complementary network, which is operative under normal circumstances.

For the maintenance of the identity of an object something else is important. The identity-maintaining constancy mechanisms operate on the neuronal representation of an object that within a temporal frame of a few seconds is extremely noisy, and this is true both for the temporal and for the spatial representation of the necessary local information. The maintenance of identity refers to the perceptual object, the special category, which is in the centre of our attention. To allow this, an abstraction must have occurred. There must exist a schema that is carried throughout some time. Without a schema, or an abstraction, maintenance of perceptual identity is not possible, because the local information both in time and space is too noisy to give reliable physical information for this task. But how are schemata neuronally represented, how can we conceive a neuronal representation of a top-down control of an abstracted image? One way to find an answer would be to look at the genesis of a percept in some detail. An essential aspect for this discussion comes from Lin Chen, who shows that topological invariants like connectivities, or the inside–outside relationship and in particular holes are recognized or perceived earlier than local features that are traditionally considered as the essential building blocks of the genesis of a visual percept. I think it is in particular important to note that ''holes'' are basic topological features, because what is an object other than a hole in the visual field?

It follows from the experiments on topological invariants that pre-semantically fragmented ''primordial objects'' predefined by the topological invariants become ''real'' objects (''what it is'' that I see) only in the second step, and this is done by linking local features to the topological invariants under the umbrella of a schema. This connection of the global topological invariants and the local signs is constructed just for a few seconds. The constancy mechanisms support for a short time the maintenance of identity, for a time in which global topological invariants and local features are bound together; ''binding'' is obviously used here with a different meaning. After a few seconds (2–3 s approximately) the identity of a perceptual object is put into question. Using metaphorical language, one could say: The brain asks in regular intervals what is new in the world, or is everything still the same? What is required is verification or falsification of what is already represented. The perceptual apparatus checks in steps of a few seconds whether there has been a change or not.

In what I have discussed, complementarity appears to be an underlying generative principle, not only a descriptive mode. My punchline is that one has to overcome the unitary principle of explanation. To understand the genesis of a percept looking only at the information flow from the periphery to some enigmatic central representation does not work as is shown by Lin Chen, but looking only for instances that fit with the Gestalt principles does not work either. I suggest that we need to seriously consider complementarity as a general frame of theoretical reasoning, and by doing so the challenge is to identify those special complementarities that are necessary to establish a visual percept. Some such candidates have been suggested: It appears to be useful to introduce the complementarity of content functions (''what it is'' that I see) and logistical support functions that are necessary for the genesis. It appears to be useful to look at local stimulus features and at global invariants and the way they are bound together within the frame of complementarity. It appears to be useful to use the principle of complementarity for stability and change of percepts, i.e., for the maintenance of identity of what is seen, and the dynamics of perception to allow the genesis of a new percept. It appears to be useful to assume complementary as it is expressed in the bottom-up and the top-down control when presemantic information has to be linked to semantically rich schemata. Finally, it appears to be useful to treat the ''what'' and the ''where'' of a visual percept within the conceptual frame of complementarity. No one entity is operative without the other; nothing can be treated without the other.

REFERENCES

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*, 704–715.

Podvigin, N. F., Jokeit, H., Pöppel, E., Chizh, A. N., & Kiselyeva, N. B. (1992). Stimulus-dependent oscillatory activity in the lateral geniculate body of the cat. *Naturwissenschaften*, *79*, 428–431.

Pöppel, E. (1988). Size constancy and oculomotor modulation of perifoveal light-difference threshold. *Naturwissenschaften*, *75*, 463–465.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, *1*, 56–61.

Pöppel, E., & Harvey, L. O. (1973). Light-difference threshold and subjective brightness in the periphery of the visual field. *Psychologische Forschung*, *36*, 145–161.

Pöppel, E., Held, R., & Frost, D. (1973). Residual visual function after brain wounds involving the central visual pathways in man. *Nature*, *243*, 295–296.

Pöppel, E., & Logothetis, N. (1986). Neuronal oscillations in the human brain: Discontinuous initiations of pursuit eye movements indicate a 30-Hz temporal framework for visual information processing. *Naturwissenschaften*, *73*, 267–268.

Pöppel, E., Schill, K., & von Steinbüchel, N. (1990). Sensory integration within temporally neutral system states: A hypothesis. *Naturwissenschaften*, *77*, 89–91.

Please address all correspondence to: Institute of Medical Psychology, Ludwig-Maximilians-Universität, Goethestrasse 31, 80336, Munich, Germany.
Email: ernst.poeppel@med.uni-muenchen.de

Beyond ''local to global'' and ''global to local''

Sheng He

Department of Psychology, University of Minnesota, Minneapolis, USA

Reading Lin Chen's target paper reminded me again how I became interested in vision research: Exciting, important, and difficult questions like the ones raised and tackled in Chen's research and summarized in the target paper.

As Chen stated, the idea that visual perception is a local to global process has dominated the field. This is not really surprising, aside from the overwhelming physiological evidence supporting the hierarchical architecture of the visual processing stream, the ''local to global'' idea fits with our naïve and intuitive conception of constructing any objects (not just visual objects). In this sense, Chen's claim to the contrary with his theory of topological perception is extraordinary. Chen's work questioned the validity of the ''local to global'' belief, and the impressive experimental evidence Chen and his colleagues have accumulated over the last 20 years has undoubtedly shaken the foundation on which we view the process of visual perception.

In my opinion, the strength of the topological theory is in its formal theoretical appeal and in the supporting psychophysical data. At this point what is not clearly established for the theory is a strong link to the known mechanisms in neurophysiology. Below I list some of the puzzles in my mind and also suggest some potential experiments, all stemming from considerations related to the neural correlates of topological perception. These issues are raised in the spirit of discussion and with the hope that Lin Chen's clarification will help me as well as others to gain a better understanding of the theory.

Linking ''local to global'' and ''global to local'' to the feedforward and feedback pathways, vs. to parallel (e.g., magnocellular and parvocellular) pathways

Because retinotopic mapping requires a rigid relationship between points in space and points on the cortex surface, by definition, neurons that follow retinotopic mapping cannot explicitly represent topological properties. Early visual cortices (V1, V2, etc.) are retinotopic, can we exclude these areas as candidates to underlie topological perception? If we put the neural correlates for topological perception following retinotopic areas (e.g., the anterior temporal lobe, as suggested by the fMRI study on long-range apparent motion from Chen's group), does that contradict the claim that topological properties are extracted earlier than other featural properties? Of course visual information processing is not a one-way street, it involves both feedforward and feedback pathways. Maybe the more invariant and global properties modulate the rigid and local processing through feedback influence. But I am not sure if mapping the local–global and global–local processing to the feedforward and feedback pathways fits with Lin Chen's original intention. In a sense, this interpretation leaves the fundamental question ''what are the primitives'' ambiguous, because then the answer to this question depends on where you are standing and which direction you are looking: Features are the primitives for feedforward processing, and global properties (such as topological properties) are the primitives for feedback processing).

Instead of a sequential or a looped configuration, another possible resolution could be that global and local properties are processed in parallel (but interactive) pathways, with the pathway for global information faster. In a sense, the parvocellular and magnocellular pathways are closest to fitting the bill, perhaps with the parvo pathway supporting the local properties and the magno pathway supporting the more global properties. Of course, this is an oversimplified correlation. However, interesting ideas could be tested related to this hypothesis: Given that the parvo pathway is the opponent pathway and the magno pathway is the broadband pathway, are we less sensitive to topological properties defined solely by chromatic differences (e.g., red hole in a green disc), presumably processed in the parvo pathway?

Perceptual saliency vs. neural processing primacy

Behavioural studies using response accuracy and/or reaction time have provided important information regarding what properties the observers are *sensitive* to. However, there is a gap between perceptual saliency or sensitivity and the primacy in neural processing. Caution has to be taken when inferring processing stages from behavioural data. More accurate and quicker response to property A than B does not necessarily mean that A is processed before B. In fact, to play the role of devil's advocate, I will present a thought experiment that illustrate the opposite.

One of the fundamental functions of vision is to perceive and recognize objects. Once we achieved the representation of objects, the features (parts) necessary for the processing are not necessarily explicitly represented. For example, when we see a smiley face, we arrive at the representation of a smiley face. In the picture, the critical piece of information for this final representation is the curved-up semicircle (mouth). However, it is possible that we will be able to distinguish between smiling and sad faces with better accuracy and shorter time than we need to distinguish between a semicircle facing up and a semicircle facing down (this is also a thought experiment). However, this result may reflect that ''face'' rather than ''semicircle'' is the final representation, and more readily available for response. Note that in this analysis, the ''semicircle'' is assumed to be processed before we reach the level of face representation yet the accuracy and response time may show that we are more sensitive to the face than the semicircle.

Is it possible then that the closer a property's representation is to the final output (the object representation), the more available that property is for guiding our intentional actions such as making a decision about the stimulus? From the perspective of invariance, at the object representation level, the representation is already independent of viewpoint, and in addition we have the well-known property of object constancy (the representation is invariant to changes of the attributes that defines the final object). At the very beginning of the input, namely the retinal image, the representation is pixel by pixel, probably the most noninvariant stage. How do we reconcile this rigid to invariant representation with the topological proposal that the invariant representation arises before the featural representations?

The role of topological properties in the action pathway

The target paper primarily focused on the perceptual measures, although the experiments on apparent motion lead one to wonder about the role of topological properties in the action pathway. If action is used as the measure (i.e., pointing, eye movements, etc.), do we expect to see even stronger indications of sensitivity to topological properties? Since a central argument of the topological theory is that topological properties are extracted early, probably at the pre-attentive stage, this implies that there will be minimal role, if any, of visual awareness in the extraction of topological properties. The honeybee study certainly supports this prediction. A somewhat related but different question, is: Do blindsight (or agnosic) patients show preserved sensitivity to topological properties?

Are topological properties adaptable?

As stated in the target paper by Lin Chen, one of the difficult problems associated with the experiments on topological perception is that any two stimuli that differ in topological properties will have to differ on other properties as well. This poses a significant challenge to the design of experiments to test sensitivity to topological properties, and forces us to rely on converging evidence from multiple experiments. However, there may be one potential solution to that problem, if topological properties are adaptable. Instead of relying on convergence across experiments, one may be able to achieve convergence within an experiment.

Adaptation is ubiquitous in the biological system. Almost all aspects of sensory input are adaptable. Examples include colour, orientation, motion direction, size, texture density, and even deviation from face prototypes. The consequences of adaptation can be seen either in the reduced sensitivity to the adapted property or sometimes more apparent aftereffects (e.g., the motion aftereffect). If topological properties are fundamental primitives that are represented early in the visual system, they of course are coded by neurons, and these neurons are highly likely to be adaptable too. For example, after an observer is exposed to many different outlined shapes (circles, rectangles, trapezoids, irregular shapes, etc., of different sizes and orientations) that all contain a dot inside (topologically equivalent), will the observer become less sensitive to the dot-inside-figure than dot-outside-figure? If so, when presented briefly with a figure that has a dot right on the boundary, will the observer more likely to report the dot is outside the figure?

Similarly, for the purpose of localizing areas sensitive to topological invariance, one may use the so-called fMRI adaptation approach: areas sensitive to topological difference, but respond similarly to pairs of different but topologically equivalent figures, should show a enhanced activation when a pair of topologically different figures are presented in sequence, but not a pair of topologically equivalent figures. One may also ask whether topological properties are represented modularly or in a distributed fashion. As mentioned above, topological properties cannot be explicitly represented in retinotopic areas, but that is based on the current conception of how vision works. Maybe that conception is itself incorrect, and there is a way to represent topological properties in the early visual cortex?

In summary, the theory of topological perception has brought a significant challenge to the traditional model of visual perception. Chen's theoretical formulation and experimental evidence forces us to consider the importance of invariance extraction in visual perception. In this regard, it is difficult to overestimate the contribution of the topological perception theory to our eventual understanding of vision. However, instead of contrasting ''local to global'' with ''global to local'', is it possible to consider visual perception as involving both ''local to global'' and ''global to local'' processes? Maybe the traditional hierarchical theory favouring local to global processing and Lin Chen's topological theory are not unlike the trichromatic and the opponent theories for colour vision.

REFERENCES

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Please address all correspondence to: Sheng He, Department of Psychology, University of Minnesota, 75 E. River Rd., Minneapolis, MN 55455, USA. Email: Sheng@umn.edu

On topology's place in the psychophysical structure of human vision

Jeremy M. Wolfe

Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, USA

Lin Chen has done us a favour by raising the visibility of topological factors in the study of perceptual organization (Chen, 1982, 1990, this issue; Zhou, Chen, & Zhang, 1992). From his vantage point, topology has a privileged position. To quote from the current target paper, he posits ''a functional hierarchy, in which the relative salience or priority in the perception of different geometric properties is remarkably consistent with the hierarchy of geometries, stratified by Klein's Erlangen program'' (Chen, this issue, p. 603). His advocacy of the central role of topology is tied to his advocacy of a ''holistic'', global-to-local approach to form perception. This he contrasts with a local-to-global approach in which we detect spots that are connected into lines that join to make angles, and so forth.

Chen's program is a useful corrective to a somnolent acceptance of a local-to-global story that we might have learned as undergraduates. However, I would argue that the claims about topology and holistic processing are too sweeping. First, the assertion of topology's primacy is too strong, at least in the realm of visual selective attention. The role of topology in early perceptual organization may be separable from its role in the guidance of visual attention. While topology may have a privileged role in the early segmentation of the image into candidate objects, it is not clear that it has similar privilege as a feature guiding selective attention. Second, Chen casts ''local-to-global'' vs. ''global-to-local'' as a two-alternative forced-choice options. I suggest that his ''global-to-local'' hypothesis is actually less holistic than some of the alternatives and, moreover, that it is not necessary to choose one account over the other. Image understanding may proceed on several paths at once.

Topology may not be pre-eminent in the guidance of visual search

Treisman introduced the idea that a limited set of features could be processed ''preattentively'', across a large portion of the visual field in parallel (Treisman & Gelade, 1980). Building on the work of others (e.g., Egeth, Virzi, & Garbart, 1984; Hoffman, 1979), I modified that idea into the concept of preattentive ''guidance'' of selective attention (Wolfe, 1994, 2003; Wolfe, Cave, & Franzel, 1989). A limited set of features could be used to direct selective attention. Thus, if an observer is looking for a red vertical item, she can use colour and orientation information to guide her attention to red items and vertical items. The intersection of the set of red items and the set of vertical items is a good place to direct your attention if your target is red vertical.

It seems intuitively clear that some features provide more robust guidance than others. Salient colour and motion are near the top of this introspective hierarchy. Properties like curvature (Wolfe, Yee, & Friedman-Hill, 1992b) or Vernier offset (Fahle, 1991), while having properties of basic guiding features, seem to have those properties to a lesser degree. I couch this discussion in very qualified language because of the difficulty in comparing across stimulus dimensions. While we can argue that the difference between vertical and horizontal is greater than the difference between vertical and 10 degrees off vertical, it is less obvious how we should compare a vertical–horizontal difference to a red–green or curved–straight difference.

Probably the best answer to this problem has come from Nothdurft's (1993) method of matching the salience of different ''pop-out'' stimuli against a common yardstick—in his case, salience in the luminance domain. Thus, it is possible to say *this* motion contrast is the same as *this* orientation contrast because they are both matched to this luminance contrast. No one has generalized Nothdurft's program to make a convincing hierarchy of feature salience. Were this done, it would be possible to say, for example, that a given topological feature and a given orientation contrast had the same salience, as matched to luminance. If it then turned out that the topological feature always produced faster or more efficient search, then the argument for topological primacy would be strengthened.

In the absence of a program of research of this sort, the examples of topological primacy can be debated without resolution. Consider Chen's Figure 22 as an example: Panel B shows a rather slow orientation segmentation compared to faster segmentation based on topological properties (holes) in panels C and D. This is a bit surprising since orientation search is typically fast and efficient (Beck, 1966a; Treisman & Gelade, 1980; Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992a). (Note: I am using ''efficient'' and ''inefficient''—rather than ''parallel'' and ''serial''—in order to describe the slopes of RT $\times$ Set size functions in visual search in a theory-neutral manner; Wolfe, 1998). Perhaps the example in Chen's Figure 22B puts orientation at a disadvantage. Symmetrical stimuli create less salient differences and less efficient searches than asymmetrical ones (Wolfe & Friedman-Hill, 1992). Moreover, one might worry that the ''holes'' in panels C and D of that figure are part of target items that are larger than the distractors, in addition to being ''holier''.

This is not to say that Chen is wrong on this point—only to note that examples can be created to argue the opposing view. Look at Figure 1. Two types of items are present: One has a hole; the other does not. In the upper panel, the orientation singleton is easy to find; in the lower panel, the topological singleton, the item with a hole, is much less salient.

It could be argued that Figure 1 has stacked the deck against holes. Figure 2 uses a very typical ‘‘O’’ vs. ‘‘C’’ hole discrimination. It is not at all clear that the hole has any special status when it comes to attracting attention. The luminance and orientation singletons seem at least as salient.

Topology might be pre-eminent in the early organization of the visual input

The figures above are not experiments. They are merely illustrations. What they illustrate is the difficulty in asserting topological primacy in the guidance of visual attention. Is there a sense in which topology does trump other features? It could be that topological constraints are key to the initial segmentation of the continuous visual input into perceptual objects. Attention in visual search tasks

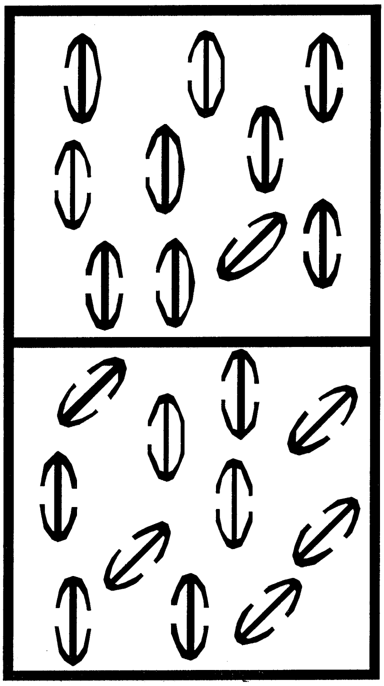

Figure 1. In the upper panel, the target is an orientation singleton; in the lower panel, it is a topological singleton. Which singleton is more salient?

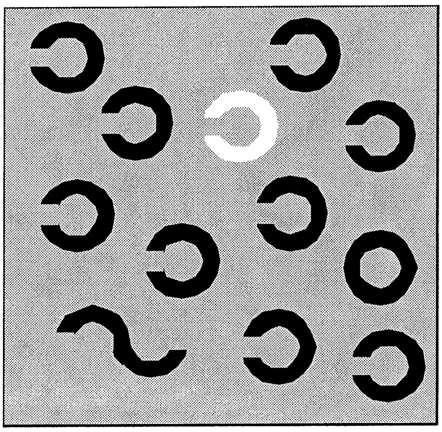

Figure 2. How do the three odd items in this figure compare in salience?

appears to be directed to objects rather than to raw features (Goldsmith, 1998; Rensink & Enns, 1995; Wolfe & Bennett, 1997), though no one really knows how to segment the visual input into these candidate objects. Search experiments suggest that these proto-objects have some structure to them. For instance, we found that properties of parts can be distinguished from properties of wholes (Bilsky & Wolfe, 1995; Wolfe, Friedman-Hill, & Bilsky, 1994) (see also Xu & Singh, 2002). It seems entirely reasonable to propose that topological constraints are vital to this first step in segmenting the image.

Dissociating early vision from attentional guidance

If topology turns out to be central to initial segmentation and if topology does not have the same privileged role in the subsequent guidance of attention, it follows that the features of early vision and those that guide attention are dissociable. Chen indirectly supports this thought when he argues against what he calls the local-to-global account of perceptual organization. The model of perception that Chen is arguing against is related to the classic two-stage architecture of Neisser (1967) and Treisman (Treisman & Gelade, 1980). A collection of local features is extracted in a first, preattentive stage and then those local features are bound into objects in a later, attention-demanding stage.

The original notion was that the set of basic feature dimensions used in early vision was the same as the set that supported effortless texture segmentation (Beck, 1966b) and efficient ''parallel'' visual search (Treisman, 1986). This seems to work, up to a point. Dimensions like colour and orientation are available early in visual processing and can be found easily in visual search tasks and will readily produce effortless texture segmentation. However, in subsequent years numerous troubling dissociations have appeared. For instance, we have shown that ''effortless texture segmentation'' can be dissociated from ''efficient visual search'' (see Figure 3 for examples based on Wolfe, 1992).

Moreover, consider line junctions (e.g., T-junctions and intersections). We know that they are available to early vision to support the division of the field into objects that can be searched (e.g., Rensink & Enns, 1998). We know they are available to attention-demanding, object-recognition processes. However, the distinction between an intersection and a T-junction is not available to guide attention (Wolfe & DiMase, 2003). A simple, two-stage linear process would have trouble explaining how critical junction information could be present, then lost, then recovered as visual information moved up a linear local-to-global sequence.

Several researchers have proposed versions of reentry or feedback to handle problems with the classic, two-stage architecture (DiLollo, Kawahara, Zuvic, & Visser, 2001; Hochstein & Ahissar, 2002). We suggest a somewhat different architecture. There is a pathway from massively parallel early visual processes to object recognition processes that match input to a multitude of possibilities in parallel but that can only handle one object at a time. An attentional bottleneck lies between early vision and object recognition. A guidance mechanism controls the selection of one object at a time. The guidance mechanism sits to one side of the main pathway. Its properties are derived from the products of early vision but they do not act as a filter between early vision and later processes like object recognition. Thus, it is possible for the guidance mechanism to lack information about a property like ''intersection'' even if the early vision and later object recognition have access to that information. In the context of the present discussion, it seems entirely possible that topology has a primary role in the creation of the primitive objects but that topological factors do not have a

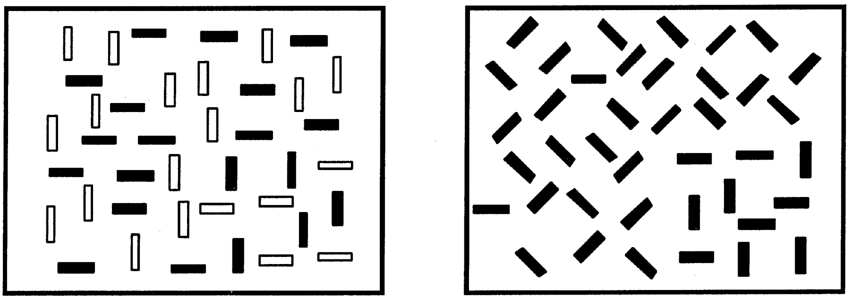

Figure 3. (a) The region defined by a different conjunction of colour and orientation does not segment but black verticals are easy to find. (b) The region defined by a different combination of orientations does segment even though isolated horizontal items are not very salient. (Adapted from Wolfe, 1992.)

similarly privileged role in guiding the selection of which of these objects will be recognized and attended.

A new experiment of Chen's may underline this dissociation. It is possible to keep track of several (four or five) identical items as they move around amidst other identical items (Pylyshyn & Storm, 1988; Yantis, 1992). Changing features that guide attention (e.g., colour) does not disrupt tracking but attacking the objectness of an object does. VanMarle and Scholl (2003) found that tracking failed when an object turned into a ''substance'' that ''poured '' from one spot to the next. Chen has done tracking experiments where the tracked objects undergo topological transformations. He found that a change in topology (e.g., a solid object developing a hole) was disruptive of the ability to track (Chen, personal communication). This seems consistent with the idea that topology may be critical in defining what counts as a perceptual object even if it is not preeminent in the guiding of attention to those objects.

The great divide?

Finally, it is worth considering Chen's ''great divide'' between local-to-global and global-to-local accounts of perceptual organization. Interestingly, it could be argued that Chen's position lies toward the middle of a continuum between the extreme positions. He rejects the notion of building up objects out of local, atomic feature elements much as the Gestalt psychologists rejected Wundt and Titchener's efforts to use introspection to reduce perception to its atoms. He argues for an account in which objects, defined initially by topological factors, are the entry-level units. Local features get attached to these objects. This is a rather moderate position compared to more radically holistic possibilities. For example, Oliva and Torralba (2001) propose a set of operators that can be used to describe scene structure and even aspects of scene meaning without ever parsing the scene into its component objects at all. They call this scene description the ''spatial envelope''. Interestingly, the operators that produce the spatial envelope are fairly simple filters that combine the outputs of different early vision units sensitive to different orientations at different spatial scales. At present, there is no definitive proof that human vision makes use of spatial envelope computations. However, models of this sort represent a holistic processing of the scene far more radical than what is proposed by Chen.

Atomic features, topological objects, holistic spatial envelopes—it is not clear that we need to assert the primacy of one of these levels of analysis. In an initial brief exposure, there is clear evidence that we are sensitive to the mean and distribution of the atomic features (Ariely, 2001; Chong & Treisman, 2003). Chen has reviewed old evidence and provided new evidence for the rapid parsing of the scene into topologically defined objects. The work on ''gist'' and picture memory suggests a very rapid ability to encode the meaning of the scene as a whole (Mandler & Ritchey, 1977; Rensink, 2000) and the spatial envelope work suggests that this ''gist'' might include a contribution from holistic processes that ignore the problem of objects altogether. It seems entirely likely that all of these analyses are activated at the onset of a new stimulus. Note that I am not arguing for an ''everything happens everywhere all the time'' view of perception. There are rules and constraints. The purpose of many of our research programs is to establish the nature of those rules. I would simply argue that it is unnecessary to propose a single path from input to image understanding. A number of parallel streams of information appear to contribute. Chen has done us the favour of underscoring the contribution of one of those streams.

REFERENCES

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Beck, J. (1966a). Effect of orientation and shape similarity on perceptual grouping. *Perception and Psychophysics*, *1*, 300–302.

Beck, J. (1966b). Perceptual grouping produced by changes in orientation and shape. *Science*, *154*, 538–540.

Bilsky, A. A., & Wolfe, J. M. (1995). Part–whole information is useful in size X size but not in orientation X orientation conjunction searches. *Perception and Psychophysics*, *57*(6), 749–760.

Chen, L. (1982). Toplogical structure in visual perception. *Science*, *218*, 699–700.

Chen, L. (1990). Holes and wholes: A reply to Rubin and Kanwisher. *Perception and Psychophysics*, *47*, 47–53.

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

DiLollo, V., Kawahara, J.-i., Zuvic, S. M., & Visser, T. A. W. (2001). The preattentive emperor has no clothes: A dynamic redressing. *Journal of Experimental Psychology: General*, *130*(3), 479–492.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 32–39.

Fahle, M. (1991). A new elementary feature of vision. *Investigative Ophthalmology and Visual Science*, *32*(7), 2151–2155.

Goldsmith, M. (1998). What's in a location? Comparing object-based and space-based models of feature integration in visual search. *Journal of Experimental Psychology: General*, *127*(2), 189–219.

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

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception and Psychophysics*, *25*, 319–327.

Mandler, J., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 386–396.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton, Century, Croft.

Nothdurft, H. C. (1993). The conspicuousness of orientation and visual motion. *Spatial Vision*, *7*(4), 341–366.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175.

Pylyshyn, Z., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, *40*(10–12), 1469–1487.

Rensink, R. A., & Enns, J. T. (1995). Pre-emption effects in visual search: Evidence for low-level grouping. *Psychological Review*, *102*(1), 101–130.

Rensink, R. A., & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research*, *38*, 2489–2505.

Treisman, A. (1986). Properties, parts, and objects. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.), *Handbook of human perception and performance* (Vol. 2, pp. 35.31–35.70). New York: John Wiley & Sons.

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

VanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects vs. substances. *Psychological Science*, *14*(5), 498–504.

Wolfe, J. M. (1992). ''Effortless'' texture segmentation and ''parallel'' visual search are *not* the same thing. *Vision Research*, *32*(4), 757–763.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*(2), 202–238.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–74). Hove, UK: Psychology Press.

Wolfe, J. M. (2003). Visual search: Are some enduring controversies moving toward solution? *Trends in Cognitive Science*, *7*(2), 70–76.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, *37*(1), 25–43.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433.

Wolfe, J. M., & DiMase, J. S. (2003). Do intersections serve as basic features in visual search? *Perception*, *32*(6), 645–656.

Wolfe, J. M., & Friedman-Hill, S. R. (1992). On the role of symmetry in visual search. *Psychological Science*, *3*(3), 194–198.

Wolfe, J. M., Friedman-Hill, S. R., & Bilsky, A. B. (1994). Parallel processing of part/whole information in visual search tasks. *Perception and Psychophysics*, *55*(5), 537–550.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992a). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 34–49.

Wolfe, J. M., Yee, A., & Friedman-Hill, S. R. (1992b). Curvature is a basic feature for visual search. *Perception*, *21*, 465–480.

Xu, Y., & Singh, M. (2002). Early computation of perceptual part structure: Evidence from visual search. *Perception and Psychophysics*, *64*(7), 1039–1054.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340.

Zhou, W., Chen, L., & Zhang, X. (1992). Topological perception: Holes in illusory conjunction and visual search. *Investigative Ophthalmology and Visual Science*, *33*(4), 958 (Abst. No. 1326).

Please address all correspondence to: J. M. Wolfe, Visual Attention Lab, Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115, USA. Email: wolfe@search.bwh.harvard.edu

Object oneness: The essence of the topological approach to perception

Jun Zhang

Department of Psychology, University of Michigan, Ann Arbor, MI, USA

How perceptual organization occurs was the central question of Gestalt psychology, and the concise mathematical articulation of the Gestalt principles remains a century-old challenge. Chen (in the target paper), in summarizing two decades of pioneering research in the topological approach to visual perception, has presented a large body of empirical evidence (using both traditional Gestalt-style experiments and modern neural imaging technology), and persuades us to think again about the psychological and computational processes underlying object perception. Chen's work has been much influenced and inspired by Gibson (1979), who advocated the importance of environment-based visual invariants and direct perception. However, Chen has gone beyond the Gibsonian proposition by bringing in formal mathematical statements to enumerate and express fundamental geometric invariants for early visual perception. The experiments in Chen (1982, 1985) supplied the empirical evidence that the topology of a stimulus configuration plays an important role in visual perception. To formalize his intuition about topological visual perception, Chen (following Zeeman, 1962) defined a tolerance relation on a discrete point set, and suggested the use of tolerance space topology to characterize global topological invariants in a visual configuration. To complete his theory, Chen (1983) proposed an information processing hierarchy for the perception of a visual figure, which involves the extraction of (in progressive order) topological invariants, projective invariants, affine invariants, and finally Euclidean invariants. This sequence coincides with Felix Klein's Erlangen program for the mathematical characterization of geometries as transformation groups acting on a space. Identifying and mapping these geometric invariants onto a perceptual hierarchy, with the perception of topological invariants at the forefront, elegantly fulfils the Gibsonian promise of a geometric theory underlying perceptual organization.

At the core of Chen's thesis is the primacy of figure–ground segregation for visual perception. To transform an image-based representation, which occurs at the retina, to an object-based representation, where figures and their background are all separated, requires the ''binding'' of points/locations into distinct regions (''chunks'') each with a certain topology. The ''glue'' that enables this binding is what motivates Chen's topological approach. Starting from a point set representing the visual input on the retina, Chen asks how to properly partition the visual space into regions/chunks solely based on large-scale topological properties.

Recall that topology on a point set deals with issues such as continuity, connectedness, neighbourhood, surroundedness, etc., and involves notions like closure, interior, boundary, etc. Insofar as objects occupy space separately from one another and from their surrounds, distinct topological relationships arise whenever there are occlusion relationships among the objects and/or between an object and its background. Take the favourite stimulus of Chen's: an object with a hole. The importance of such a stimulus is that, depending on whether the boundary of the hole belongs (or is perceived as belonging) to the exterior as opposed to the interior part of the ''hole'', one's percept switches from that of a doughnut (with a hollow centre that unveils its background) to that of a solid disk (in front of a continuous background). This type of ownership of boundary/border is, in set-theoretic language, a question of whether or not a set is defined to include its boundary ∂, i.e. whether the set is closed or open. This is clearly a distinction at the level of topology.

Chen's topological approach advocates a global-to-local order of processing, in the sense that processing of an object's topological property takes precedence over identifying an object's features—that is to say, the establishment of object as a whole, or object ''oneness'', precedes the identification of specific features belonging to an object. Chen's proposal is provocative, yet carefully reasoned. His information processing hierarchy placed the extraction of topological (and Euclidean) invariants as the first (and last) step. Though it might appear counterintuitive since one would expect the Euclidean properties of a visual image to be recorded right at the outset, Chen argued that Euclidean invariants, such as rigidity (invariant under mental translation and rotation), really are tag-on properties of an already-segregated visual object and are therefore computed after a stimulus is treated as a topological whole. This view challenges traditional computer vision algorithms, where object segregation is based on the identification and binding of features.

Chen's topological proposal has far-reaching consequences for computational algorithms of object oneness, and thus is worth scrutinizing. The following comments will examine Chen's specific suggested use of tolerance space topology on a discrete set, and discuss an alternative approach based on the topology of continuous spaces, i.e., a topological manifold. Moving from a discrete to a continuous setting allows one to conveniently impose differentiability conditions, thereby turning a topological manifold into a differentiable one with a fibre bundle structure. The two central concerns from Chen's topological visual perception, namely the characterization of object oneness and the characterization of shape-changing transformations, will be shown to admit a natural interpretation under the fibre bundle/Riemannian manifold model of visual perception.

Tolerance topology on discrete sets

Chen, following Zeeman's (1962) influential paper, proposed to use a particular type of discrete topology, called the tolerance space topology, to characterize the global topological properties of objects. A tolerance on a point set is a binary relation (i.e., among any two elements/points of the set) that is both reflexive and symmetric, but not necessarily transitive. The absence of a requirement for transitivity makes a tolerance relation different from an equivalence relation. This is an important distinction, because equivalence (reflexive, symmetric, and transitive) relations are the starting point for many common topics of topology, such as the quotient operation; in order to obtain nontrivial global topological properties on discrete sets, one is forced to use this tolerance relation (in lieu of the equivalence relation) to represent perceptual ''indistinguishability''.

Despite it being a topology on a discrete set, tolerance topology allows the definition of paths, connectedness, holes, and dimensionality. The space of tolerance relations, the tolerance space, is identified as the mathematical characterization of the stimulus configuration. Chen then invokes the algebraic topological notion of homotopy group, first suggested in the context of visual perception by Zeeman (1962) and Zeeman and Buneman (1968), to characterize the tolerance structure among the stimulus points. Specifically, a simplicial complex (i.e., a complex made of simplexes) can be constructed with its vertex points being the points in the original point set. Edges connect pairs of points that are within a given tolerance. For three distinct points, if all pairwise distances are within the tolerance, they form a triangular face; accordingly, they are indistinguishable from one another under this tolerance. The same holds for four, five, ... distinct points, and so on. In this model, distinguishability is characterized by missing edges, faces, etc., in the higher dimensional simplexes that make up the complex. When embedded into the Euclidean space, the dimensionality of the complex increases linearly with the total number of points in the stimulus configuration. Though not a problem in principle, the structure of this simplicial complex, and the resulting homology group H, may become extremely complicated and difficult to compute except for very few dots in the configuration. It remains a challenge to demonstrate that, with the criteria for spatial tolerance becoming either more relaxed or more stringent, a change of H would parallel the change in the resulting percept. In short, Chen's tolerance topology relies on the fundamental assumption of discreteness of visual stimulus configuration as input to vision perception. Even though it may appear as a simplifying assumption, the discrete set approach may turn out to suffer severe computational disadvantages.

Topological manifold and fibre bundle

An alternative to the tolerance topology idea of Chen (borrowed from Zeeman) is to introduce a manifold structure on visual space, so that visual perception takes place on a topological manifold. A (topological) manifold is formed by continuously pasting together pieces of Euclidean space. The only requirement imposed on the point set is Hausdorff separability, namely, for any two distinct points, there exist disjoint open sets that each point is contained by. This continuity property about visual inputs allows one to set up Cartesian coordinate systems (called ‘‘charts’’) at each point on the manifold, so that neighbouring points can be specified using these coordinates. Different charts centred on the same point are related to each other via a coordinate transformation. When a certain collection of charts covers the entire manifold, it is called an ‘‘atlas’’.

Topological manifold captures the basic architecture of information processing by the visual system: Neurons earlier in visual processing stream respond to inputs from restricted regions of the visual space, and that the entire visual space is covered by the overlapping receptive fields (charts) of the neuronal ensemble. In order to compare and contrast features extracted from nearby points, one needs to provide for a proper calculus on the topological manifold. This is achieved by supplying additional (differentiable) structure to make a topological manifold a differentiable one. One may, on a differentiable manifold, perform covariant (intrinsic) comparisons of vectors located at neighbouring points, accomplished through a geometric entity called an affine connection. If the manifold is further endowed with a metric tensor, then it becomes a Riemannian manifold, which admits a unique (called Levi-Civita) connection.

The argument that visual perception involves a stimulus manifold describable in terms of a Riemannian manifold has traditionally appeared in the study of binocular space and depth perception (for example, Indow, 1982, 1991; Luneburg, 1947; Smith, 1959; Yamazaki, 1987). The idea of stimulus comparison in multidimensional perceptual space using covariant differentiation was also explored in Levine (2000). However, none were addressing the issue of object oneness, namely the binding of contiguous locations on the base manifold into a topological whole. In a radical departure from these traditional approaches, Zhang and Wu (1990) used Riemannian geometry to characterize neural processes mediating the segregation of figure–ground relationships and the topological layout of the visual space. Based on identifying the tangent space of the visual manifold as that of motion (directional) selective neuronal responses, and that object oneness is reflected as the intrinsic constancy (through parallel transport) of tangent vectors across neighbouring points, the Levi-Civita connection Γ of the visual manifold is established. Solving for the metric tensor g yields the following expression of the Riemannian metric:

$$\mathbf{g} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix}$$

where f denotes the grey-level intensity function (of a two-dimensional image denoted by spatial coordinates x,y) and that the subscripts denote respective partial derivatives. So any image f induces, through its second derivative (Hessian), a metric tensor and a resulting Levi-Civita connection. It was proven (Zhang & Wu, 1990) that the Riemann-Christoffel curvature R of this connection Γ is identically zero, so parallel translations of a vector are path-independent—this in turn means that segregation of image regions (objects) is possible globally. Specifically, motion-based object segregation, in which image points ''glue'' together if they are a part of an object undergoing rigid translation (i.e., with spatially uniform image velocity), is represented as a region at which the tangent vectors are intrinsically constant (with vanishing covariant derivative).

Take the example of a random-dot kinematogram (Braddick, 1974), which highlights, on the one hand, the remarkable ease at which object oneness is established by our visual system and, on the other hand, the difficulty with any computation algorithm of object segregation based on feature analysis, as forcefully argued by Chen in the target paper. The image luminance of successive frames allows the motion system (directional sensitive neurons) to extract local features in terms of local movement directions. However, because of the aperture problem, the local directions and the direction of the global target displacement would not agree. Furthermore, because the background dots were also being randomly displaced, motion sensors respond to these regions as well, resulting in a nonuniform response map by motion sensors (Zhang, 1995). This nonuniform response map, or tangent vector field V, is to be compared and construed under the Riemannian metric g. Global topological properties are extracted by covariant differentiation of V (motion response map). The advantage of this Riemannian geometric framework is that the chicken-and-egg problem of whether to compute features or objects first is avoided—objects defined by the constancy of their physical features (e.g., velocity) across space necessarily give rise to an intrinsically constant vector field under an affine connection and, therefore, can be immediately segregated using geodesic coordinates (see Zhang, 1995 for more details). The emergence of a visual figure (target) is the result of simultaneously solving the aperture problem and the location-binding problem.

Though constructed in a continuous (rather than discrete) setting, the differential manifold (fibre bundle) model of visual perception resonates with Chen's basic argument about the primacy of spatial proximity in establishing object oneness—his idea that proximity takes precedence over similarity. This is because, in the language of differential manifold, proximity is simply the (geodesic) distance between points on the base manifold while similarity/dissimilarity is represented by the covariant difference of vectors (i.e., visual features) situated on different base points of the stimulus manifold. The former involves a unique, metric-compatible Levi-Civita connection while the latter may use any affine connection defined on the appropriate fibre. Chen's ideas about proximity taking precedence over similarity precisely expressed the distinction between points on the base manifold (related by proximity) and points in the feature space (related by similarity).

Characterizing topological deformation of an image

One of the questions raised by Chen's topological approach to visual perception is about the characterization of rubber-sheet (plastic) deformations of objects in an image—Chen referred to them as ''shape-changing transformations''. Correspondence of an object across different images, e.g., in apparent motion, may be established even when the object undergoes considerable deformation. Our visual system's ability to detect and recognize the same object despite a topological deformation (with limited extent) is often called ''shape constancy''. Though intuitively easy to describe, the precise manner and degree of visual distortion of an object, however, is hard to quantify mathematically based on computation of grey-level image properties alone. Previously, Leyton (1992) systematically investigated the underlying general linear transformation group and demonstrated how a combination of ''stretch'', ''shear'', and ''rotation'' operations (which form appropriate subgroups themselves) on the object's symmetric axis would result in different shapes that nevertheless would be recognized as being produced by the same object. However, none of these operators were generated by specific images themselves, and therefore given an arbitrary image, one does not know which operators to apply and what symmetric axes are appropriate at each image location. One needs a set of image descriptors or curvilinear image coordinates that these shape-changing transformation groups can apply *locally*.

One such descriptor was provided in Zhang (1994). It was based on computing the second derivative (Hessian) of the image function f. More precisely, the eigen-vectors of the image Hessian are computed at each image location and, assuming their smoothness, a flow field can be constructed using either of the eigen-directions. These two orthogonal flow fields will fill up a patch of the two-dimensional visual manifold; together they become the curvilinear coordinates that capture local invariant structure of the image function. An image is allowed to deform along either coordinate curves (i.e., the value of a pixel may be dragged by those flows). To avoid the problem of noncommutativity of the two directional vector fields, Zhang (1994) reparameterized the flow fields to make them bona-fide (i.e., mutually compatible) coordinate curves; this was done through forcing their Lie bracket operator to commute, a necessarily condition for orthogonal flow fields to be orthogonal coordinate curves. The infinitesimal transformation of a visual contour, embodied as the Lie derivative dragging the flow field along its path, the so-called ''orbit'' of a Lie group, quantifies topological transformations such that the invariance (''psychological constancy'') of a contour under the transformation group is reflected as its being annulled by the action of the Lie derivative (Hoffman, 1966). The only freedom remaining, the so-called ‘‘gauge freedom’’, is with respect to the scaling of these image-dependent coordinates; this flexibility is important because we want the amount of deformation to have some arbitrary scales. Examples of selecting (i.e., fixing) a particular gauge for ‘‘good’’ or Gestalt images were presented in that paper—it turns out that the original Cartesian space where the image function is defined and the curvilinear coordinates where deformation is quantified are related through a conformal transformation. This computational theory and the associated algorithm for characterizing shape-changing transformation closely follow the spirit of the Lie Transformation Group (LTG) approach proposed by Hoffman (1966, 1968, 1970, 1989, 1994). While it is in no sense complete, hopefully it is a first step towards finding a representation of rubber-sheet deformation (of an image) that is parameterized by the image itself.

Conclusion

To summarize, Chen's research in topological visual perception forces the computational vision community to rethink the difficult problem of object oneness. As Chen cited ‘‘Everything is difficult at its very beginning’’; it is particularly true if this beginning involves specifying a proper topology for visual perception. Whether to use the tolerance topology on discrete sets or topological (and differentiable) manifold of fibre bundles, future research will clarify the most suitable topological framework to precisely capture the notion of object oneness.

REFERENCES

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research*, *14*, 519–527. Chen, L. (1982). Topological structure in visual perception. *Science*, *218*, 699–700.

Chen, L. (1983). What are the units of figure perceptual representation. *Studies of Cognitive Science, No. 22*. School of Social Science, University of California, Irvine).

Chen, L. (1985). Topological structure in the perception of apparent motion. *Perception*, *14*, 197–208.

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton-Mifflin.

Hoffman, W. C. (1966). The Lie algebra of visual perception. *Journal of Mathematical Psychology*, *3*, 65–98. Errata *4*, 348–349.

Hoffman, W. C. (1968). The neuron as a Lie group germ and a Lie product. *Quarterly of Applied Mathematics*, *25*, 423–441.

Hoffman, W. C. (1970). Higher visual perception as prolongations of the basic Lie transformation group. *Mathematical Biosciences*, *6*, 437–471.

Hoffman, W. C. (1989). The visual cortex is a contact bundle. *Applied Mathematics and Computation*, *32*, 137–167.

Hoffman, W. C. (1994). Conformal structures in perceptual psychology. *Spatial Vision*, *8*, 19–31.

Indow, T. (1982). An approach to geometry of visual space with no a priori mapping functions: Multidimensional mapping according to Riemannian metrics. *Journal of Mathematical Psychology*, *26*, 204–236.

Indow, T. (1991). A critical review of Luneburg's model with regard to global structure of visual space. *Psychological Review*, *98*, 430–453.

Levine, D. N. (2000). A differential geometric description of the relationships among perceptions. *Journal of Mathematical Psychology*, *44*, 241–284.

Leyton, M. (1992). *Symmetry, causality and mind*. Cambridge, MA: MIT Press.

Luneburg, R. K. (1947). *Mathematical analysis of binocular vision*. Princeton, NJ: Princeton University Press.

Smith, A. A. (1959). The geometry of visual space. *Psychological Review*, *66*, 334–337.

Yamazaki, T. (1987). Non-Riemannian approach to geometry of visual space: An application of affinely connected geometry to visual alleys and horopter. *Journal of Mathematical Psychology*, *31*, 270–298.

Zeeman, E. C. (1962). The topology of the brain and visual perception. In M. K. Ford (Ed.), *Topology of 3-manifolds and related topics* (pp. 240–256). Englewood Cliff, NJ: Prentice-Hall.

Zeeman, E. C., & Buneman, O. P. (1968). Tolerant spaces and the brain. In C. H. Waddington (Ed.), *Theoretical Biology: Vol. 1. Prolegomena* (pp. 140–151). Birmingham, AL: Aldine Publishing Co.

Zhang, J. (1994). Image representation using affine covariant coordinates. In Y.-L. O, A. Toet, D. Foster, H. J. A. M. Heijmans, & P. Meer (Eds.), *Shape in picture: Mathematical description of shape in grey-level images* (NATO ASI Series F: Computer and Systems Sciences, Vol. 126, pp. 353–362). Berlin: Springer-Verlag.

Zhang, J. (1995). Motion detectors and motion segregation. *Spatial Vision*, *9*, 261–273.

Zhang, J., & Wu, S. (1990). Structure of visual perception. *Proceedings of National Academy of Sciences USA*, *87*, 7819–7823.

Please address all correspondence to: Jun Zhang, Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA. Email: junz@umich.edu