Spatial Vision, Vol. 8, No. 4, pp. 383–391 (1994)
© VSP 1994.

Theoretical issues in symmetry perception

CHRISTOPHER W. TYLER
Smith-Kettlewell Eye Research Institute, 2232 Webster St., San Francisco, CA 94115, USA

> O Nature, and O soul of Man! how far beyond all utterance are your linked analogies! not the smallest atom stirs or lives in matter, but has its cunning duplicate in mind.
>
> Herman Melville, *Moby Dick* (1850, p. 295).

This Feature Issue is the first of a pair devoted to the topic of the human perception of symmetry relations in the visual world. At a time of increasing *rapprochement* between the analyses of local spatial properties and the global organization of Gestalt pattern relations, symmetry relations can provide a guiding principle to help connect the two. It therefore seems a good juncture to take stock of research in symmetry perception, which began over a century ago with the trenchant analysis of Mach (1886). The pace of investigation in the field has been accelerating over the past two decades, although this seems to be the first time a set of articles on the topic has been collected together in one publication. This first issue addresses predominantly theoretical aspects of symmetry perception, while the subsequent issue will focus on more empirical questions.

The intent of the collection is provide insight into new analyses that are being developed for the perception of symmetries of various types. The main kinds of symmetry of interest in spatial vision are reflection (or bilateral), rotation (or axial), translational (or repetition) and size-scaling (or fractal self-similarity), many of which are explored in the following papers. Taken together, the contributions to this Feature Issue, as diverse geographically as they are conceptually, reveal a multiplicity of perspectives on the processing of symmetry by the human brain.

NEURAL REPRESENTATION OF SYMMETRY

It is hard to think about the perception of symmetry without considering the nature of its neural representation. How are physical object relations, such as their inherent symmetries, represented in the brain? This raises the classic chestnut, 'To what degree is the neural representation isomorphic with the physical properties of the object and to what degree is it abstracted?'

The simplistic answer of the Gestalt school was that representation is isomorphic with the stimulus, so that a symmetric object would have a symmetric representation in the brain. Although now generally dismissed as begging the question of the representation code, this view has been resuscitated by Shepard (1981), who argues that an isomorphic representation is necessary to allow internal transformations of the representation to match manipulations of the object in the world. This stress on a veridical representation is an elaboration of Young's (1962) paradoxical insight that higher organisms maintain their difference from the environment only by *mirroring* the properties of the environment in the brain. This perspective emphasizes the structural symmetry between subject and object in the act of perception. The experiments of Shepard and colleagues on such transformations as mental rotation of random objects seem to provide good evidence for a 3D neural representation of the 3D world. Indeed, the classic experiments by Penfield (1959) of the sequential readout of memory sequences, and our mental ability to replay songs and speeches in temporal order, would extend the isomorphic representation to the fourth dimension of time.

However, the problem of a representation code remains. The presence of an isomorphic copy of a relevant object in the brain may have value, but it is still subject to the criticism that the object has not been encoded into some form that captures the connotations of its features to the organism. Minsky (1975) and Pylyshyn (1986), for example, argue for an entirely propositional code for object features with no coherent spatial representation. Such a code would resemble lists of attributes associated with each object, where each item would be cross-referenced to other related properties that would constitute its meaning to the organism. This view of encoding arises from the position that much of our sense of meaning must be carried by such a propositional or cognitive code, so it is parsimonious to assume that perception is similarly coded. For symmetry, such a code would consist of a list of attributes for the base pattern motif and a specification of the transformations required to generate the complete pattern from operations on its base motif. Without such a list of transformations, the propositional code could not be said to have encoded the symmetry of the pattern.

Shepard has argued cogently that a propositional code alone is insufficient to account for many of the properties of object recognition under spatial transformations. In particular, the speed of matching objects that are rotated copies of each other is proportional to the angle of rotation but independent of the complexity of the objects. Neither result would be predicted from the inherent properties of a propositional code, but both are consistent with the idea of a neural representation of the objects in which elements that are close in physical space are represented as close in the neural connectivity space. The comparison of objects at different angles of rotation would then correspond to a matching process after the appropriate transformation has been applied to the neural representation. This transformation would correspond not to a physical rotation in the brain but to an adjustment of the local neural codes for each part of the object equivalent to such a rotation (as opposed to merely updating the propositional code for the transformation itself). One advantage of an isomorphic code is that it allows operations such as filtering, segmentation and spatial relations to be performed in a natural way by local neural operations. What is known of the neurophysiology supports the idea of an isomorphic mapping of 2D space to retinal space, of depth via binocular disparity and of time via velocity coding (DeYoe and van Essen, 1988).

However, the issue of how symmetry is encoded neatly underlines the limitations of an isomorphic code, which would require that the object be translated, rotated, dilated or otherwise transformed in all possible combinations for the symmetry properties to be discovered. Given the time established for mental rotation to take place (e.g. Cooper and Shepard, 1973, 1978), the isomorphic hypothesis would require that it would take many minutes to become aware of the possible structural symmetries of even 2D images, whereas these relations seem to be immediately evident for brief presentations of less than 100 ms (Corballis and Roldan, 1974; Hogben *et al.*, 1976). The implication is that there must be some means of encoding the similarity relations implicit in the similarity structure in a more direct manner than is available by an isomorphic representation. Is this the point of transition to a propositional code, or is there some more geometric code for these fundamentally geometric relationships?

PROPERTIES OF A PERCEPTUAL ENCODING HIERARCHY

The preceding considerations lead to the view that there are four basic aspects of perceptual encoding forming the structure within which any perceptual task, such as symmetry perception, should be viewed.

Encoding the metric of the stimulus space

The first property of perceptual encoding is the metric within which the referent space is represented in the brain. In some cases, such as retinal space, the metric is a relatively faithful match to the stimulus space. In others, such as color, it may be drastically reduced in comparison to the available stimulus metric. Some examples of the metrics of perceptually relevant spaces are:

a 1D metric for time,
a 2D metric for color (hue and saturation),
a 3D metric for space,
a 6D metric for rigid objects in space (three dimensions of position and three dimensions of rotation around each position).

The encoding metric may fall far short of the stimulus metric in dimensionality; for example, a random dot field has an N-dimensional metric, where N is the number of dots. We may encode these as mostly indiscriminable random textures, an example of failing to distinguish the metric dimensions themselves rather than just the positions along the dimension. Only idiosyncratic examples may evoke sufficient encoding response to form distinguishable dimensions. A less extreme example of encoding insufficiency is the dimensionality of perceived rotations. Intuitively, it is obvious that there are three dimensions of positional variation, but most people need to try it out to convince themselves that rotation around a point also has a dimensionality of three.

In recognition of their imperfect reflexivity, Shepard (1981) refers to the match between the encoding metric and the stimulus metric as one of complementarity, the relationship between a lock and key or a photographic negative and its print. Piaget (1969) originated a similar concept in referring to schemata, by which the organism accommodates itself to the nature of the object (in order to assimilate the object into the arrangement desired by the organism). The schema is a broader concept than Shepard's complementarity, including action components as well as the representative aspects, but it carries the full sense of the reciprocity between the mental representation and the world that is to be brought under its hegemony.

The structure of the representation in the coding metric

After establishment of the internal space of the coding metric, the next question is the structure of the representation in each dimension of the metric. Is the metric a continuous one, like an intensity code along the metric dimension, or are there discrete channels with local preferences? This distinction has been emphasized by Foster (1982), although he included structured representations of higher-order features in the discrete category, whereas here such higher-order features are considered to form dimensions of the coding metric itself. This distinction relates to the question of whether each dimension is intensive (coded by the intensity of a single channel) or extensive (coded by an array of similar channels differing in one respect). But this is a subsidiary question that has little effect on the structure of the encoding metric.

Note that the presence of independent channels at one level of encoding may be converted into a continuous representation by a push-pull, or opponent, linkage between channels, as in the classic example of the opponent mechanisms of color vision. This conversion highlights the properties of the encoding mechanism in determining the encoding metric. The presence of more than one channel whose outputs make essentially independent contributions to the response results in a discrete channel code; linkage between channels to determine the output generates a continuous code for each pair of linked channels (assuming a graded response for each channel alone). Another way to obtain an essentially continuous code is to have a large number of independent channels across the dimension, forming a continuous extensive coding through the channel space as opposed to the continuous intensive coding of the push-pull arrangement.

If there is a discrete channel coding, a key question to answer is how many channels lie along the coding dimension (in each local retinal region). Many techniques have been developed to address this question but few of them have been worked out in sufficient detail to have full confidence in the answers (Tyler *et al.*, 1993). Historically, channel modeling in vision began with discrete channel models, as exemplified by the threshold elevation paradigm developed in color vision by Stiles (1939, 1959). Discrete channel analysis in spatial vision goes back to Wilson and Bergen (1979), followed by Wilson *et al.* (1983), Swanson and Wilson (1985) and Foley and Yang (1991), among many others. It also has been used for a variety of other stimulus domains, such as temporal frequency (Mandler and Makous, 1984; Anderson and Burr, 1985; Hess and Snowden, 1992) and stereomotion (Beverley and Regan, 1973).

A subsequent question, when the channel distribution on each dimension has been established, is the nature of the interactions between channels. Strictly speaking, a channel is an entity that is independent of its companion channels, so that interacting channels constitute an overall channel. However, the situation is analogous to that of neural receptive fields, where the excitatory region of the receptive field is considered to be the primary or 'classical' receptive field. Surrounding regions that modulate the response of the excitatory region without generating a response when stimulated by themselves form the integration field that presumably represents interactions (excitatory or inhibitory) with adjacent neurons (McIlwain, 1964; Fischer and Kruger, 1974; DeAngelis *et al.*, 1992). In the same way, paradigms may be designed to distinguish between primary and interactive aspects of channel behavior. Color provides a canonical example of such a distinction, where the cone sensitivity functions for intensity increments may be determined by appropriate isolation techniques but these primary channels then interact in opponent fashion to provide chromatic information. As long as there is empirical access to both levels of organization, there is no categorical problem in distinguishing between the primary and interactive aspects of the channel structure.

The more concrete aspect of encoding structure, its physiological instantiation in the hardware of the brain, is tangential to the present discussion of its logical structure. For example, realizing an intensity code as a neural firing rate adds little to our appreciation of the properties of the code; it could equally well be a cellular voltage or a concentration of transmitter molecules, which indeed seem largely interchangeable with the firing rate code in the operation of neurons. Rather, it is the functional organization of the encoding process that determines its effectiveness to the organism.

The coding of image symmetries by a self-matching or autocorrelation process

Beyond the low-level structure of the representation, there arises the issue of the encoding of regularities or symmetries in the image by capitalizing on their redundancy to simplify the representation. These are issues that have been raised variously by Gibson (1950), Garner (1962) and Attneave (1954), among many venerable figures. The mechanisms for implementing this simplification remain unresolved, however.

A variety of schemes has been proposed to address the issue of how symmetries may be encoded. One of the most sophisticated is that of Palmer (1982), a transformational code based on the fact that the global structure of the coding space is determined by the symmetry relations in the local operators (cortical filters) that are generating the coding space. Thus, a symmetry relation that exists in the image will evoke multiple matching patterns in the response space. This operation reduces the complexity of the problem because all types of symmetry in the image space are converted simply to one type of symmetry in the coding space: translational symmetry. However, as Palmer points out, a second-order comparison mechanism is required to compare the outputs in the coding space and determine whether such matches exist.

The need for comparison stages of second (and higher) order reveals that it is the connectivity relations in the comparison stage that determine what aspects of symmetry in the image are processed. In fact, Palmer postulates that the higher-order stages are local within the coding space, which means that global symmetries would be detected only by their local symmetry relations. If we had only local connectivity, for example, we could not detect that the ears are symmetrically located in a face unless we had detectors large enough to encompass both ears across a facial image.

Rather than postulating second-, third-, and higher-order matching mechanisms for detecting symmetry relations (as does Palmer, 1982), one may postulate a general-purpose autocorrelation mechanism operating in the coding space to detect symmetries of any type at any range. Tyler and Miller (1994) have shown how the core element of repeated or translationally symmetric patterns may emerge through an autocorrelation process of mutual reinforcement. One way that such a mechanism might be implemented in the nervous system is for the local pattern vector at the focus of attention (usually the fovea) to form a template that is correlated automatically with the local pattern vector at all other locations. By 'local pattern vector' is meant the profile of response strength through the array of local detectors of all sizes, orientations and types. A high correlation of this vector between two locations means that the same pattern of inputs was present at the two locations at some orientation and scale. Any mechanism that could make such comparisons therefore would act as a detector of all the types of symmetry present in each local pattern vector.

The specific representation of the object in the coding space

Objects are represented by clusters of features with specifiable symmetry relations, but there is a kind of perceptual 'surface tension' that holds the features of a specific object together in the representation space. There must be a vocabulary of object forms or schemata of the type proposed by Piaget (1969) to which are attached the features specific to the object in the current focus of attention. The concept of a schema is understood as an action structure; e.g. a 'box' is a cuboid structure with an inside in which other objects may be placed. Part of this action structure is its three-dimensional symmetry relations, i.e. the aspects or regions of the object that may be made congruent by a symmetry transformation. Unlike the 2D symmetry structure (which, as argued in the previous section, is likely to be coded by inherent symmetries in the coding space), the complexity of generalized 3D symmetries probably require internal manipulation of the type proposed by Shepard (1981) to ascertain its symmetry structure.

I would suggest that the total process of object characterization may involve both structural and manipulative aspects in the symmetry encoding. For example, if a cube is viewed in perspective projection with one side facing the observer, the quartic symmetry of the front face might be perceived by inherent coding symmetries, implying that it does not need to be mentally rotated around the line of sight for the quartic symmetry to be appreciated. Conversely, the symmetrical identity of the six faces (or lack of identity if the cube is a rectangular cuboid), can be appreciated only by mental rotation of the front face into the other positions. In this way, the structural aspect of the symmetry coding process may be brought to bear in those regions of the image where manipulation is not required; the results of the structural analysis may then be carried into the manipulative phase so that features that are found to be congruent by a manipulative analysis consequently are perceived to have the same inherent symmetry structure.

There may also be a level of cognitive assumptions in the manipulative aspect of the encoding that, for symmetry operations, takes the form of avoiding the need for some mental rotations on the basis of an assumed symmetry relation. For example, when looking at a set of banisters in perspective projection, one may need only to perform the mental operation of matching one banister to the next to establish the similarity and then make the cognitive assumption that all the others are similarly matched. This assumption would economize on the number of mental manipulations that need to be made at the cost of missing deviations from the assumed symmetry in regions where manipulative testing is avoided.

Much of what we see may be only part of the total object schema; e.g. when sitting on a couch, we may have only an arm and part of the seat cushions projecting on to the retina, but perceive them as part of the complete couch. This process of amodal (or non-sensory) completion has been most fully explored by Kanisza (1976), although he emphasized regions hidden by other objects rather than regions outside the visual field. The concept of an abstract schema representing the whole object gives a direct instantiation to the process of amodal completion; if part of an object is sufficient to evoke unambiguously the full schema, the whole object is perceived in the implied position. This evocation is similar to the process known as 'lexical access' in language perception (Marslen-Wilson and Tyler, 1982), in which the full meaning of a spoken word is evoked as soon as it is unambiguously distinguishable from all other words. In the same way, the whole object may be perceived when the information from the visible parts is sufficient to identify it from the vocabulary of known objects or plausible object classes. There is the obvious proviso that the manner in which the non-visible parts are obscured must be consistent with their occlusion by an intervening object of some kind, so that attended region of the scene is itself a plausible arrangement. That there are counterexamples in which the attended region makes an 'impossible' object (Hochberg, 1978) does not reduce the general applicability of this proviso.

CONCLUSION

The four processing stages above constitute a framework in which most studies of perception in general and symmetry perception in particular may be viewed. One aspect that may be considered to be understressed is the role of attention in this framework, but this is in accord with the view of perception as a largely parallel, subconscious process for which attention is the mechanism of access or gateway to the higher, serial processes of cognition. This gateway presumably occurs somewhere in the fourth of the stages described, the object representation stage, and channels the preprocessed symmetry information to consciousness to aid the organism in dealing with its environment.

In terms of their predominant position in this encoding scheme, the papers by **Joseph** and **Victor** and by **Matin** and **Li** belong perhaps at the initial level of the encoding structure. Each presents a novel form of perceptual stimulus constraint whose perceptual structure is to be determined, the former for scale of random-element patterns and the latter for the direction of perceived vertical. The second level is perhaps the home for the analysis of symmetry-based receptive-field structure by **Kurbat**, and will be of more relevance for the subsequent Issue on Empirical Studies in Symmetry Perception. A group of papers focus on questions relating to the third level, the perceptual processing of symmetry redundancies in visual images. Papers by **Dakin** and **Watt**, by **Zabrosky** and **Algom**, by **Latimer**, **Juong** and **Stevens**, and by **Bonneh**, **Reisfeld** and **Yeshurun** each propose a different computational scheme for extracting 2D symmetry information and compare it to results from human observers on the same tasks. Spanning between the third and fourth levels of encoding, the papers by **Vetter** and **Poggio** and by **Pani** address in different ways the use of symmetry relations in 3D objects to extract their 3D structure.

Acknowledgement

This work was supported by NIMH grant MH 49044.

REFERENCES

Anderson, S. J. and Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Res.* **25**, 1147–1154.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193.

Beverley, K. I. and Regan, D. (1973). Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. *J. Physiol.* **235**, 17–29.

Cooper, L. A. and Shepard, R. N. (1973). The time required to prepare for a rotated stimulus. *Memory and Cognit.* **1**, 246–250.

Cooper, L. A. and Shepard, R. N. (1978). Transformations on representations of objects in space. In: *Handbook of Perception* (Vol. 8). E. C. Carterette and M. P. Friedman (Eds). Academic Press, New York.

Corballis, M. C. and Roldan, C. E. (1974). On the perception of symmetrical and repeated patterns. *Percept. Psychophys.* **16**, 136–142.

DeAngelis, G. C., Robson, J. G., Ohzawa, I. and Freeman, R. D. (1992). Organization of suppression in receptive fields of neurons in cat visual cortex. *J. Neurophysiol.* **68**, 144–163.

DeYoe, E. A. and van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci.* **11**, 219–226.

Fischer, B. and Kruger, J. (1974). The shift effect in the cat's lateral geniculate nucleus. *Exp. Brain Res.* **21**, 225–227.

Foley, J. M. and Yang, Y. (1991). Forward pattern masking: effects of spatial frequency and contrast. *J. Opt. Soc. Am.* **8**, 2026–2037.

Foster, D. H. (1982). Analysis of discrete internal representations of visual pattern stimuli. In: *Organization and Representation in Perception.* J. Beck (Ed.). Lawrence Erlbaum, Hillsdale, NJ, pp. 319–341.

Gibson, J. J. (1950). *The Perception of the Visual World.* Houghton Mifflin, Boston.

Garner, W. (1962). *Uncertainty and Structure as Psychological Concepts.* Wiley, New York.

Hess, R. F. and Snowden, R. J. (1992). Temporal properties of human visual filters: Number, shapes and spatial covariation. *Vision Res.* **32**, 47–59.

Hochberg, J. (1978). *Perception.* Prentice-Hall, Englewood Cliffs, NJ.

Hogben, J. H., Julesz, B. and Ross, J. (1976). Short-term memory for symmetry. *Vision Res.* **16**, 861–866.

Kanisza, G. (1976). Subjective contours. *Scient. Amer.* **234**, 48–52.

Mach, E. (1886). *The Analysis of Sensations.* Dover, New York.

McIlwain, J. T. (1964). Receptive fields of optic tract axons and lateral geniculate cells: Peripheral extent and barbiturate sensitivity. *J. Neurophysiol.* **27**, 1154–1173.

Mandler, M. B. and Makous, W. (1984). A three channel model of temporal frequency perception. *Vision Res.* **24**, 1881–1887.

Marslen-Wilson, W. and Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition* **8**, 1–71.

Minsky, M. (1975). A framework for representing knowledge. In: *The Psychology of Computer Vision.* P. H. Winston (Ed.). McGraw-Hill, New York.

Palmer, S. E. (1982). Symmetry, transformation and the structure of perceptual systems. In: *Organization and Representation in Perception.* J. Beck (Ed.). Lawrence Erlbaum, Hillsdale, NJ, pp. 95–144.

Penfield, W. (1959). The interpretative cortex. *Science* **129**, 1719–1725.

Piaget, J. (1969). *The Mechanisms of Perception.* G. N. Geagrin (Trans.). Basic Books, New York.

Pylyshyn, Z. (1986). *Computation and Cognition.* MIT Press, Cambridge MA.

Shepard, R. N. (1981). Psychophysical complementarity. In: *Perceptual Organization.* M. Kubovy and J. R. Pomerantz (Eds). Lawrence Erlbaum, Hillsdale NJ, pp. 279–342.

Stiles, W. S. (1939). The directional sensitivity of the retina and the spectral sensitivities of the rods and cones. *Proc. Roy. Soc. Lond.* **B 127**, 64–105.

Stiles, W. S. (1959). Colour vision: The approach through increment threshold sensitivity. *Proc. Natl. Acad. Sci.* **45**, 100–114.

Swanson, W. H. and Wilson, H. R. (1985). Eccentricity dependence of contrast matching and oblique masking. *Vision Res.* **25**, 1285–1295.

Tyler, C. W., Barghout, L. and Kontsevich, L. L. (1993). Computational reconstruction of the mechanims of human stereopsis. In: *Computational Vision Based on Neurobiology.* T. B. Lawton (Ed.). *Proceedings of the SPIE* **2054**, 52–68.

Tyler, C. W. and Miller, R. T. (1994). Pattern identification by trajectory analysis in autocorrelation hyperspace. *Proceedings of the World Congress on Neural Networks* **III**, 312–316.

Wilson, H. R. and Bergen, J. (1979). A four-mechanism model for threshold spatial vision. *Vision Res.* **19**, 19–32.

Wilson, H. R., McFarlane, D. K. and Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Res.* **23**, 873–882.

Young, J. Z. (1962). Why do we have two brains? In: *Interhemispheric Relations and Cerebral Dominance.* V. B. Mountcastle (Ed.). Johns Hopkins University Press, Baltimore, MD.