Title
Transformational Structure and Perceptual Organization

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 3(0)

Author
Palmer, Stephen E.

Publication Date
1981

Peer reviewed

TRANSFORMATIONAL STRUCTURE AND PERCEPTUAL ORGANIZATION

Stephen E. Palmer
University of California, Berkeley

The problems of structure and organization in perception are among the most central in the history of the field. Beginning with the gestalt movement early in this century, researchers and theorists alike have supposed that certain types of perceptual phenomena provide insights into the structure and organization of the system that produces them. There are two major questions. First, what is the common element in these phenomena that provides the most transparent view of the internal structure of the system? Second, how is this structure embodied and used within the system so that it produces the observed phenomena?

In this paper I suggest answers to both questions. To provide an initial idea of the direction I will take, here is a brief preview. The organizational phenomena I take to be most critical for understanding the structure of perception are: (a) object constancy despite changes in image size, shape, position, and orientation, (b) motion perception of structured objects, (c) figural goodness or "good gestalt", (d) perceptual grouping of the visual field, and (e) reference frame effects. First, I will argue that these phenomena are related by the fact that underlying them all is a common transformational structure. Second, I will describe a general design for a computational system within which transformational structure can be analyzed easily. The proposal is to couple a transformationally based system of analyzers with an attentional mechanism that establishes a variable frame of reference within it. The job of the attentional frame is to move around within the space of analyzers so as to maximize the invariance and stability of the attended output. It does so by compensating for the changes brought about by the transformations. Finally, I will make a few remarks about why I think the kind of computational structure I am proposing is an interesting one from a purely theoretical standpoint.

Transformational Structure

The basis of transformational structure is the concept of transformational invariance. Transformational invariance refers to the fact that when an object undergoes a spatial transformation, such as a rotation, a great many changes occur in the pattern of stimulation on the retina, but at the same time there is a great deal of higher order structure that does not change at all. All sorts of relationships (or relationships among relationships) do not change, even though each first order property does, and these unchanging aspects are the transformational invariants. In many cases they correspond more directly to the intrinsic properties of the real world object undergoing the transformation -- its size, shape, color, and so forth -- than to the properties of its projected image. The image, after all, changes in ways that the object clearly does not. Because transformational invariants reflect object properties rather than image properties, computing them probably plays an important role in getting from an image based representation to an object or world based one.

How can the transformational structure of an event be computed from the spatio-temporal images that arise from it. The problem, of course, is that while it is trivial to compute the image over time from knowledge of the real world object and its real world transformation, the reverse is not at all trivial. In fact, there isn't even a unique solution, but only an infinite set of object/transformation pairs. In other words, there are many different world events -- objects undergoing transformations -- that could give rise to the same specific event images and no logical basis on which to decide the correct one from purely optical information. It is clear that some additional assumptions are needed about either the nature of the object, the nature of the transformation, or both to reach a determinate solution. To reach the <u>correct</u> solution as well, the additional assumptions will have to be chosen in a principled way.

What assumptions might help here? Clearly a good bet would be any assumption that is generally true of events in the world. Then, arriving at solutions that are consistent with these assumptions will almost always result in veridical perception.

Perhaps the most striking fact about the transformations that characterize events in the world is that both objects and transformations tend to be fairly stable over time and space. This is certainly true over short intervals of time and small regions of space and is usually true over long intervals of time (on the order of at least whole seconds) and larger regions of space as well. To the extent that this is so, objects can be well approximated as rigid and the transformations they undergo as uniform motions in three dimensional space. Naturally, there are many cases of non-rigid objects undergoing various complex motions. But even these are probably best understood in terms of how they can be analyzed into a structure of roughly rigid components undergoing roughly uniform motions. The motion of a person walking is a case of non-rigid motion that has yielded to such an analysis (Johansson, 1973; Cutting, 1981).

The rigidity and uniformity assumptions suggest that the perceptual system operates in such a way as to <u>maximize</u> invariance in both objects and motions. In anthropomorphic terms, the system "wants" to analyze both objects and motions as "changing as little as possible". "Wanting to change as little as possible" refers simultaneously to the facts that that the perceived object/motion pair must, in some sense, account for the sensed variations in the image and that this can be done in more than one way. To pick a well studied example, the kinetic depth effect, if a line changes simultaneously in its projected length and orientation, it might be either (a) a line that gets shorter and longer by certain amounts as it rotates in varous possible ways or (b) a line of constant length that is rotating in depth (Wallach & O'Connell, 1953). The latter is almost invariably perceived, although it sometimes takes an observer several seconds to achieve it. Once it is perceived, however, it is stable and does not spontaneously change to something else. According to the present line of thought, the preferred interpretation arises because it is "simpler" than the alternatives given the heuristic assumption that objects in the world tend to be rigid and undergo uniform motions in three dimensional space. Thus, the perceptual system seems to prefer interpretations of greatest possible invariance.

Motion and Constancy

The perceptual phenomena most directly and obviously related to transformational invariance are those of motion perception and object constancy. Motion occurs when the position of some object or part of an object changes over time. It is the paradigmatic case of a transformation, of course, but it is perhaps not so obvious what it has to do with invariance. In fact, the whole concept of a distinct object undergoing motion presupposes invariance in that the object is taken to be unchanging (except for its position, of course) as it moves. Logically, one could just as well say that the world had changed its intrinsic nature over time. This would be a more reasonable notion if one perceived the visible surfaces of the world like a rubber sheet that simply changed its shape plastically during events. The fact is that people to not perceive the world in this way, but as consisting of articulated, unchanging objects that undergo various sorts of motions. This highlights the fact that perceiving motions of objects actually presupposes invariant aspects as well as varying ones, and that an event in the world always has both components.

It appears, then, that object constancy is just the other side of the coin from motion perception. Motion is the perceived transformation; object constancy is the perceived invariance. They are completely coupled in that for the motion to be different, the object must be different too. In the case of a rotation in depth, for example, either the object is rigid and the motion is uniform (as in actual depth rotation) or the object is plastic and the motion is nonuniform in such a way that their combined changes produce the two dimensional image (as in the plastically deforming colored regions in a motion picture of a depth rotation).

Given that real world events tend to consist of rigid objects in uniform motions, a perceptual system would be biased toward veridicality if it somehow embodied preferences toward perceiving rigid objects and uniform motions. Indeed, there is good evidence that this is true. When presented with ambiguous information, people tend to perceive rigid objects rotating or translating in three-dimensional space as long as the optical structure is consistent with such an interpretation and the system is given sufficient time. For example, Johansson (1950) showed that people have a strong tendency to see two moving points as fixed at the ends of a rigid rod moving in three dimensions rather than as moving non-rigidly in two dimensions. Thus, an ambiguous object in unambiguous motion tends to be perceived as rigid rather than plastically deforming. The other side of this story is that an unambiguously specified object in ambiguous stroboscopic motion tends to be seen in uniform motion (Shepard & Judd, 1976; Farrell & Shepard, 1981). Clearly, there is something special about rigid objects and motions for the human visual system.

Figural Goodness

Another important problem in perceptual theory that is intimately related to transformational structure is what psychologists have come to call "figural goodness." Figural goodness refers primarily to subjective feelings of order, regularity, and simplicity in certain figures as opposed to others. The relation of this subjective feeling to transformational structure is not intuitively obvious, but it is nevertheless simple to grasp.

Figures are "good" to the extent that they are themselves invariant over certain types of transformations. The most obvious case is that of standard bilateral or reflectional symmetry. To illustrate, the letters "A" and "M" are reflectionally symmetric about their vertical axes because each letter is the same as itself after being reflected about a vertical line through its center. The other widely known type of symmetry is rotational. The letters "N" and "Z" are rotationally symmetric through an angle of 180-degrees because each letter is the same as itself after being rotated by 180-degrees. Still other letters have transformational invariance over a number of different transformations: "X", "O", "H", and "I" all have two reflectional symmetries (about vertical and horizontal lines through their centers) as well as 180-degree rotational symmetry. A perfect circle has still greater transformational invariance because it is unchanged by all central rotations and reflections.

There are two other, less well known types of symmetry: translational and dilational (Weyl, 1952). They are defined by the same abstract scheme as for reflectional and rotational symmetries. Both of these latter sorts of symmetries technically apply only to idealized, infinite patterns, but one can define "local" versions that apply to finite patterns by only requiring invariance for part of the pattern over the transformation. (See Palmer, in press, for a more complete discussion of symmetry, local symmetry, and their relation to transformational structure.)

It turns out that the goodness of figures can be well predicted from its symmetries in this extended sense: the set of transformations over which the figure is invariant. Garner (1974) showed that ratings of perceived goodness increased monotonically with the number of transformations over which the figure is invariant. Garner actually talks about "rotation and reflection (or R & R) subsets" of a figure, but this concept turns out to be isomorphic to the number of rotational and reflectional invariants (Palmer, in press). Further, the amount of transformational invariance a figure has also strongly affects how quickly people can match two figures for physical identity, how well they remember a figure, and how easily they can describe it. Such results demonstrate the reality of figural goodness in perceptual processing. (See Garner, 1974, for a review). There is additional evidence that figural goodness depends on translational and dilational invariances as well (Leeuwenberg, 1971).

In summary, figural goodness seems to be characterized quite nicely by the concept of transformational invariance. The relevant transformations in this case are reflections, rotations, translations, and dilations. Except for the addition of reflections, these are the same set that characterized motion and object constancy phenomena. It seems unlikely that this is merely a coincidence.

Grouping

The next phenomenon I want to relate to transformational structure is grouping (Wertheimer, 1923). Figure 1 shows some standard examples of grouping phenomena in which most people report perceiving either a vertical or horizontal organization. Figure 1A demonstrates the influence of proximity on grouping. The dots are organized into vertical columns (rather than horizontal or diagonal rows) because their vertical proximity is greater than their horizontal proximity. Figure 1B demonstrates the influence of similarity of orientation. All else being equal, similar elements tend to be grouped together and dissimilar elements grouped apart. Many different kinds of similarity have been shown to affect grouping, but color, size, and orientation are particularly striking. Continuity, symmetry, and closure are three other well documented factors. But perhaps the most potent of all is what gestaltists called "common fate." Elements group together by common fate when they move in the same direction at the same rate. Even a completely homogeneous field of random dot texture spontaneously organizes into figure and ground when a spatial subset of the dots begins to move together or when the rest of the dots begin to move around them.

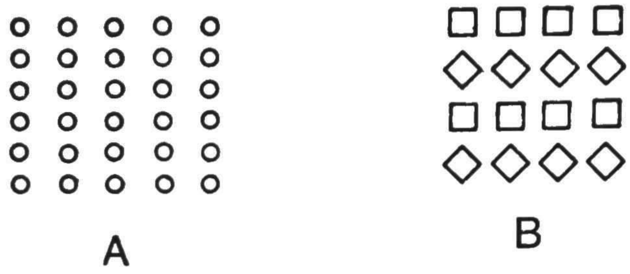

Fig. 1: Grouping by proximity and similarity.

The point I want to make about grouping phenomena is just this: Elements are grouped together when they are in closer transformational relationships to each other than they are to other elements. Transformational "closeness" refers to the magnitude of the transformation required to achieve transformational invariance. For example, the dots in Figure 1A must undergo a larger translation to bring them into congruence with their vertical neighbors than with their horizontal neighbors. In Figure 1B, the figures must undergo a rotation as well as a translation to coincide with their vertical neighbors, whereas just a translation will suffice for the horozontal neighbors.

Now consider some of the most potent factors in grouping phenomena. Similarity of two elements in position, orientation, and size can be defined by the magnitudes of the transformations -- translations, rotations, and dilations, respectively -- that are required to make them equivalent. Continuity is similarity over local translations, and bilateral symmetry is similarity over reflections. And so, once again, we find the same types of transformations lurking behind grouping phenomena as we found behind motion, object constancy, and figural goodness. Grouping seems to be determined by maximizing transformational relatedness within a perceptual group.

Frames of Reference

The final category of perceptual phenomena I want to discusss I will call "reference frame effects." It includes a number of different results in many different domains. What they all have in common is to suggest that perception at any moment occurs within a single, unitary frame of reference that captures common properties of the whole display. Other properties are perceived relative to this frame, very likely in terms of deviations from it.

As an example, consider how the orientation of a global reference frame can affect shape perception. Figures 2A and 2B show the same form in two orientations that differ by a 45-degree rotation. Figure 2A is perceived as a square because its sides are horizontal and vertical, and Figure 2B is perceived as a diamond because its sides are diagonal. However, an interesting thing happens when a number of such forms are aligned diagonally. The perceived shapes reverse: horizontal and vertical sides produce the appearance of diamonds while diagonal sides produce the appearance of squares (Attneave, 1968; Palmer, in preparation). It seems that the tilt of the whole configuration is somehow "factored out" of the display, and the orientation of the sides is then perceived relative to the whole configuration. The conjecture is that this "factoring out" is done by establishing a tilted frame for the figure within which 45 degrees is the referent orientation. This would explain why the shapes are perceived as they are, and it fits well with many other orientational phenomena in shape perception. (See Rock, 1973, for a review).

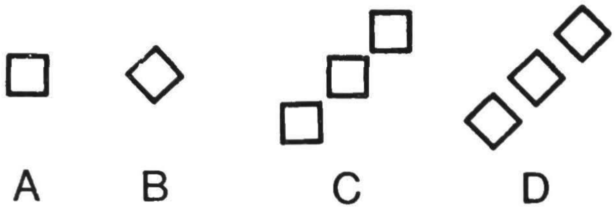

Fig. 2: Reference frames in shape perception.

Similar effects are well known to occur in motion perception. For instance, when two points are in sinusoidal motion, one vertically and the other horizontally as shown in Figure 3A, people do not usually perceive them as such. Rather, the see a configuration that moves diagonally as a unit, within which the two dots move toward and away from each other as depicted in Figure 3B (Johansson, 1950). Here again, it seems that the perceptual system establishes a frame of reference for the common motion and "factors it out." "Induced" motion effects are similar in that an unmoving object is seen to move because a larger, more prominent optical structure serves as the frame of reference, but is moving so slowly that its motion is not detected (Dunker, 1929).

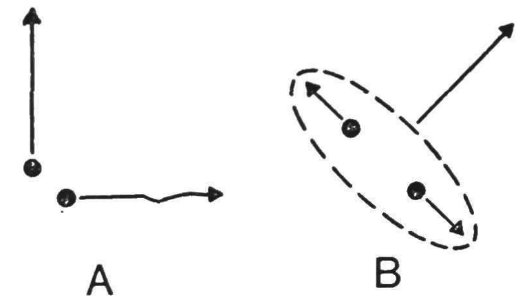

Fig. 3: Reference frames in motion perception

The relation of reference frames to transformational structure can best be explained by analogy to their role in analytic geometry. There, a reference frame is established to describe a point's location numerically. It provides a set of assumptions within which spatial positions maps into numerical coordinates. Different reference frames map the same point into different coordinates, but they are related by exactly the same transformations that relate their corresponding frames. In standard Euclidean geometry, this set consists of translations, rotations, dilations, and reflections. Thus, the structure underlying all possible Euclidean frames of reference is based on transformational invariance: the set of transformations that make one reference frame equivalent to all others.

Returning to perception, it seems that the perceptual orientation of a line or perceived direction of motion is like the position of a point in analytic geometry: their values depend on the reference frames within which they are perceived. Like its geometrical counterpart, this frame seems to include something akin to a position (origin), orientation (axis), resolution (unit size), and reflection (sense). And if this frame is variable from one moment to the next, then an underlying structure of transformations is implied that relate one frame to another. These transformations are exactly the same as we have encountered repeatedly: translations (for position), rotations (for orientation), dilations (for unit size), and reflections (for sense).

Transformational Theory of Perceptual Structure

All of the perceptual phenomena just discussed suggest that the perceptual system has a definite preference for processing optical structure involving certain kinds of transformational invariances. The questions I now want to address are (a) what this might tell us about perceptual organization and (b) how such a system might be constructed computationally.

I think the examples considered above are telling us that the visual system is built on a transformational base that can be used to extract transformations as a heuristic for solving perceptual problems. In other words, the system is designed to be transparent to transformations of the sort most often encountered so that it "prefers" interpretations involving them. By "transparent" I mean that (a) these transformations can be computed rapidly and easily in such a way that (b) the system can compensate for them simply and efficiently. This strongly suggests that the system must be designed to solve the problems of transformational invariance right from the start, and that these design features form the heart of the system.

There are three basic components in the solution I will consider here: (a) a space of analyzers that are transformationally related to one another, (b) higher order analyzers that compute output similarity of lower order analyzers over local transformational relations, and (c) an attentional mechanism that establishes a perceptual reference frame within the analyzer space that maximizes invariances. The output of attentional fixations is stored in memory as a representation of the perceived object. I will discuss each part in turn more fully, but the reader should remember that they are interrelated proposals that only make sense within the complete systemic structure.

First-order Analyzer Space

The first component consists of a set (or sets) of analyzers computing spatial properties of the visual field in parallel. Surprisingly, almost any sort of analyzers will do, as long as they have particular structural relationships to each other. The structural constraint is that they be transformationally related to each other. Precisely what this means is developed more fully and formally elsewhere (Palmer, in press), but the basic notion is just this: Two analyzers are transformationally related if their "receptive fields" or "spatial functions" are identical except for a transformation from a specific set. In the present case, the set consists, not surprisingly, of the transformations discussed earlier: translations, rotations, reflections, and dilations (the so-called "similarity" transformations of Euclidean geometry). I call such sets of transformationally related analyzers functional systems because they compute, in this transformational sense, the same spatial function (Palmer, in press).

An example would be a set of excitatory "bar detectors" with inhibitory surrounds (ala Hubel & Wiesel, 1962) whose elements differ only in the position, orientation, and size of their receptive fields. Each bar detector is related to each other one by a translation, rotation, dilation, or some composite of two or more of these transformations. A similarly constructed set of edge detectors would constitute another functional system, distinct from the first because there is no transformation from the specified set that makes a bar-like receptive field into an edge-like receptive field and vice versa.

The overall structure of a functional system can be conceptualized as an analyzer space in which each analyzer is a point. The dimensions of the space correspond to the transformational relations among them -- i.e., the position, orientation, resolution (or size), and reflection (or sense) of the analyzers relative to each other. Since the number of analyzers is certainly finite, the space is only sparsely populated with analyzer points. Therefore, it is more appropriate to think of it as something like a discrete lattice structure such as depicted in Figure 4. The cyclic dimension is orientation (which repeats after 180-degrees of rotation) and the binary dimension is reflection (which repeats after each reflection), while both positional and resolutional dimensions are simple orderings. The diagram is naturally a simplification of the actual space, since one cannot depict a structure of more than three dimensions in real space. To think of the whole structure in concrete terms, one can conceive of the vertical dimension of the structure shown in Figure 4 as resolution and then imagine a whole two dimensional array of them (to represent the two positional dimensions) like a case full of beer cans. Notice that the relations between pairs of analyzers in the space reflect the transformational relations between them as discussed above. This transformational structure defines the system.

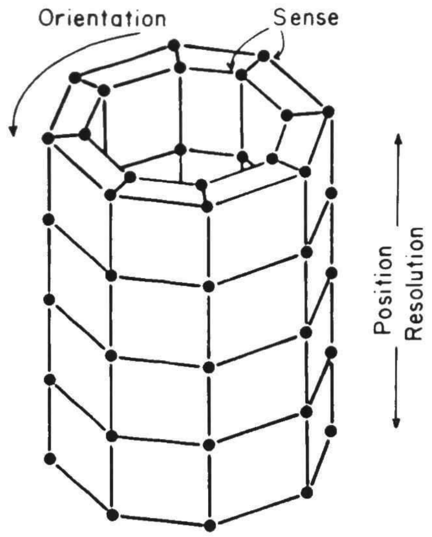

Fig. 4: The space of analyzers.

Higher-order Analyzers

The importance of transformational relatedness among analyzers is that their output is guaranteed to be the same given any two patterns (or portions of patterns) that are identical over the transformation that relates them. For instance, consider a pattern with reflectional symmetry about a vertical axis such as the letter "A". As discussed before, this means that it is invariant over a reflection about this vertical line. Now consider any analyzer that covers any portion of this pattern. Its output, whatever that might be, must be identical to that of the other analyzer related to it by reflection in the same vertical line. Moreover, this is generally true for all pairs of analyzers related to each other by this particular transformation given any pattern having that type of symmetry. Thus, the interesting fact about transformational relatedness among first order analyzers is that transformational regularities in the stimulus event will be reflected in easily computable regularities in their outputs.

In general, higher order analyzers are elements that compute such relationships among the outputs of lower order analyzers. They respond to symmetries or motions, depending on whether they compare outputs simultaneously or over a time lag. We have just discussed a case involving symmetry, but it turns out that motion analysis has exactly the same logical structure except for the introduction of a temporal difference. For example, suppose that a pattern at some time produces outputs in the first order analyzers and that the pattern then moves, say, by a translation. The output this produces after a short duration will be exactly the same as the output it produced initially for each pair of analyzers related by that translation over that time lag.

These second order analyzers can be conceptualized as the links in the lattice structure depicted in Figure 4, although there probably would be far more of them. They represent local transformational relations among first order analyzers. They compute regularities in space (symmetry) or space-time (motion) by determining transformational invariances. One can even think of them as being embedded in the same space as the first order analyzers because they have the same sort of transformational structure. In the case of motion analyzers, of course, there is the additonal dimension of rate of motion.

This transformational structure of the second order analyzers allows the possibility of piggy-backing still higher order analyzers on top. That is, the outputs of second order analyzers could be compared for similarity in just the same way that they compare the outputs of first order analyzers. This makes most sense for motion analyzers. Those analyzers that compare outputs simultaneously would provide information about symmetries and regularities of motion. Those that compare over a further time lag would provide information about accelerations and decelerations.

Visual Attention and the Mind's Eye

Given such a space of analyzers, how can it be used, as intended, to "factor out" transformational structure? The real problem here is to find some internal transformation that will compensate for the external transformation. One transformation "compensates" for another if applying the second after the first yields the identity transformation -- i.e., invariance or no change at all. For example, suppose an object is coming directly toward an observer. Over time, the image of this object expands uniformly in the visual field. If the observer were to move away from the object at the same rate as the object moved toward the observer, then the two transformations will exactly cancel, and the image of the object will not change. Thus, the "moving backward" transformation by the observer exactly compensates for the object motion.

I want to suggest that something similar happens inside the head. Rather than the eye compensating for transformations by moving about in the world, however, I suggest that visual attention moves about within the analyzer space, playing the role of the mind's eye. Like the eye with respect to the world, visual attention can change its position and orientation with respect to the analyzer space. Unlike the eye, it accomplishes both of these transformations by simple movements within the analyzer space. That is, rotations of the mind's eye correspond to translations of visual attention along the orientational dimension of the analyzer space. Similarly, changes of scale ("zooming" ala Kosslyn, 1981) can be accomplished by translations along the resolution dimension.

Perceptual Reference Frames

At the heart of this proposal is the hypothesis that visual attention is localized within the analyzer space and is centered at a particular position. This establishes a perceptual reference frame for further perceptual analysis. Fixing visual attention on one position of the analyzer space induces a reference frame because it determines a position (or origin), orientation (or axis), direction along that orientation (or sense), and resolution (or unit of distance) relative to which the contents of visual attention are coded. Thus, positioning visual attention determines the description given to objects under analysis. This is entirely analogous to the role of reference frames in analytic geometry. A circle has a different equation when the origin of its reference frame is at its center than when the origin is off-center. The corresponding phenomenon in perception occurs when the same figure looks like a square or a diamond, depending on the orientation of the reference frame within which it is perceived (see Figure 2).

If visual attention acts as a perceptual reference frame for constructing desriptions of shapes, then it is clear that certain rigid transformations can be completely compensated for by corresponding attentional changes within the analyzer space. The result is analogous to a geometrical reference frame displacing as the circle did or changing its scale size as the circle grew larger so that its equation did not change despite the changes in the geometric figure. It is easy to see that the changes in reference frame that would accompany displacements of such an attentional mechanism would be able to compensate for rigid transformations of objects, keeping the contents of the attentional frame -- whatever they might be -- constant despite the transformation. It is not difficult to imagine that such a system would "prefer" to register uniform transformations of rigid objects rather than complex motions of deforming objects. Compensating for motions of rigid objects can be accomplished simply by an attentional transformation and it does not require any change in the description of the object. Any other sort of transformation requires changes in the object's description as well. If the attentional frame somehow manages to follow the path within the analyzers space that maximizes object constancy and motional uniformity, then its operation will embody such "preferences."

Attentional Control

This leads directly to the next problem: how this attentional reference frame is controlled. It should be mentioned at the outset that there is a certain amount of conscious control over the attentional reference frame. People usually can, if pressed, attend to specified positions, sizes, and orientations. But I suspect that conscious control of attention is a high level cognitive activity that does not usually extend down to the level at which we are currently dealing. Rather, it seems that a great deal of the nitty-gritty details of attentional control must be strongly determined by stimulus structure.

How might this be done? The answer I want to explore is that the structure of stimuli determines how visual attention is positioned -- their symmetries and regularities. The basic idea is that attention is positioned to maximize transformational invariance. This can be done by finding the maximal output from the higher order analyzers for a given region of the analyzer space, since these analyzers are the ones sensitive to transformational invariance.

It is important to realize here that different reference frames imply different symmetries and regularities. Therefore, any "economy of coding" scheme for representation (e.g., Attneave, 1954) requies that the reference frame be chosen to maximize such symmetries. To illustrate this fact, consider again the case of a circle. When the origin of the reference frame is at its center, the circle has all possible reflectional and rotational symmetries centered about that point. When the origin is off-center, its only symmetry about that point is a single reflection about the line joining it to the circle's center. Such facts will be represented in the outputs of the second order analyzers at the center and off-center positions within the analyzer space. The output will be much greater at the position within the analyzer space that corresponds to the center. Therefore, there will be a strong tendency to establish the attentional reference frame at the center of the circle.

For a circle, there will be no particular orientation preference, precisely because it has complete central symmetry. For a square, however, or for any other figure that has significant asymmetries, there will be decided preferences in orientation. A square is symmetric about only the lines joining opposite midpoints of its sides or opposite vertices of its angles. Therefore, only these four orientations are serious candidates. If the midpoint line is used, the figure would have a different perceived shape than if the vertex line were used. In fact, these two frame orientations result in the "square" and "diamond" interpretations, respectively. Obviously, not all figures have exact symmetries like circles and squares do. But the same principles would apply to approximate symmetries.

The notion I have in mind for the placement of attention is a "hill climbing" process. Its goal is to maximize transformational invariance in the stimulus information by seeking the position of highest output in the analyzer space, at least locally. Sometimes there will be several maxima, and in these cases quite different percepts will arise when different maximal positions are chosen for the perceptual reference frame. The square/diamond and ambiguous triangles are two well known examples (Attneave, 1968; Palmer, 1980; Palmer & Bucher, 1981).

A single attentional fixation will seldom be sufficient to code a whole scene or complex object. More complete coding would be accomplished by making many attentional fixations at other positions within the analyzer space that have high ouputs. There is evidence of a bias toward beginning at the most global level (Navon, 1977). A reasonable guess would be that after one or two global fixations of an object at a low resolution level of the analyzer space, many local fixations would be made at higher resolution to code details. I am assuming that the contents of these attentional fixations are somehow stored in memory to form a representation of the world. The result will be a hierarchical structural description, much like those I have discussed previously (Palmer, 1975, 1977).

Given that attention can be positioned to maximize transformational invariance, it is not difficult for it to continue to do so if the object begins to move. That is merely a matter of tracking the same maximum through the analyzer space with the help of the motion analyzers discussed earlier. Recall that these analyzers are sensitive to transformational invariance over time, and, therefore, that maintaining maximum transformational invariance will entail following the maximum output of these analyzers. Doing so has the effect of maintaining object constancy over the transformation. For example, when a square begins to rotate, it is perceived as such, not as a square that changes in perceived shape until it becomes a diamond and then changes back into a square again. The latter is what would be expected if the rotation were not followed by the reference frame initially used to code its shape, but were fixed in the same unchanging orientation. I suspect that the latter is what happens when people are shown a tight spiral pattern rotating, yet see it as circles contracting into the center.

Organizational Phenomena Revisited

We now begin to see how motion can be analyzed and constancy can be maintained within such a transformationally based system. The transformations are initially coded by the higher order analyzers and then used to achieve and maintain maximal constancy. The motion finally perceived is not a simple function of the motion analyzers, since it too depends on a reference frame that maximizes invariance. This is accomplished by an attentional mechanism that finds and follows maximal output levels within the analyzers sensitive to motion and their higher-order analyzers.

Transformations of this reference frame can compensate for stimulus transformations, thereby maintaining constancy. It seems necessary that much of this must be done outside conscious attention, however, because there hardly would be enough of it to go around. More likely, once an object's representation has been established in memory, its representational schema can follow the approriate reference frame without conscious attention. This monitoring process would only require conscious attention if something unexpected were to occur, such as the object disappearing or changing its intrinsic properties.

Many perceptual grouping phenomena are a natural result of this attentional process working within a transformationally structured space. It seeks the maximal amount of transformational invariance at different levels of resolution and codes elements together that are closely related within the analyzer space. Given that attention can only cover a portion of the analyzer space and that it is attracted to local maxima, it will tend to code together items that are transformationally similar.

Finally, reference frame effects result from an attentional mechanism that is centered on a position within the analyzer space and a coding scheme, more fully specified elsewhere (Palmer, in press), that describes shape relative to the reference values of the frame. The fact that frame effects generally show that global structure affects local structure more strongly than vice versa suggests that global information tends to dominate in determining the position of the attentional frame. Higher order structure of the whole configuration seems to strongly affect the placement of the reference frame and, therefore, to influence the resulting perception.

In all of these phenomena it is clear that the system's preference for invariance over transformations holds within the three dimensional space of the world. It may hold in the two dimensional space of images as well, but when the two conflict, the simpler three dimensional solution generally dominates. We have been discussing the analysis of two dimensional images, and it is not entriely clear how to extend the proposal into the third dimension. One possibility would be to add second, three dimensional level that embodied the same design features, but in a higher dimensionality. Another would be to translate the relevant simple transformations in three dimensions into their complex counterparts in two dimensions. I do not yet have a well defined proposal to make on this difficult issue.

<u>The Importance of Systemic Structure</u>

Before closing, I want to say a few words about an initeresting property of the theory quite apart from its ability (or lack thereof) to account for perceptual phenomena. I am intrigued by it <u>systemic</u> nature. The reader may have noticed that in explaining the theory I made almost no reference to the specific nature of the analyzers that comprise the pieces of the system. At one point I said that they might be something like bar- or edge-detectors, but that was merely for illustration. In fact, it makes very little difference what the analyzers look like as long as they satisfy certain symmetry conditions: namely, they cannot be symmetrical about any transformation proposed to exist within the analyzer space. The reason should be clear. If the analyzers are invariant over a transformation, then that transformation cannot exist as a dimension within the analyzer space. For example, if all the analyzers were rotationally symmetric, as are circular center-surround receptive fields, then the system could not support the orientation dimension or the higher order rotational motion analyzers.

In any case, the fact that the constraints on the system are so weak suggests that it is primarily the structure of the whole system that is doing the work. Indeed, this must be true if the basic building blocks of the system are transformational relations. Transformational relations are an emergent property of systems of analyzers; they are simply undefined for any individual analyzer without a systemic context of other analyzers. This suggests that the internal structure of individual analyzers might best be considered in terms of their functional role within the system. Further, the nature of the elements of the system might actually be determined by optimization of their functional roles within the system as a whole (Palmer, in press). I think these are interesting and important notions for perceptual theory. They hark back to the gestalt claim that emergent properties of whole systems play the critical role in understanding perceptual phenomena. Perhaps they were right.

REFERENCES

Attneave, F. Some informational aspects of visual perception. <u>Psychological Review</u>, 1954, <u>61</u>, 183-193.

Attneave, F. Triangles as ambiguous figures. <u>American Journal of Psychology</u>, 1968, <u>81</u>, 447-453.

Cutting, J. E. Coding theory adapted to gait perception. <u>Journal of Experimental Psychology: Human perception & Performance</u>, 1981, <u>7</u>, 71-87.

Dunker, K. Ueber induzierte Bewegung. <u>Psychologishe Forschung</u>, 1929, 12.

Farrell, J. E., & Shepard, R. N. Shape, orientation, and apparent rotational motion. <u>Journal of Experimental Psychology: Human Perception & Performance</u>, 1981, <u>7</u>, 477-486.

Garner, W. R. <u>The processing of information and structure</u>. Potomac, Md.: Erlbaum, 1974.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction, and the functional architecture of the cat's visual cortex. <u>Journal of Physiology (London)</u>, 1962, 160, 106-154.

Johansson, G. <u>Configuration in event perception</u>. Stockholm: Almqvist & Wiksell, 1950.

Johansson, G. Visual perception of biological motion and a model for its analysis. <u>Perception and Psychophysics</u>, 1973, <u>14</u>, 201-211.

Kosslyn, S. M. Image and mind. Cambridge, Mass.: Harvard University Press, 1981.

Leeuwenberg, E. L. J. A perceptual coding language for visual and auditory patterns. American Journal of Psychology, 1971, 84, 307-350.

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

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. Hillsdale, N.J.: Erlbaum, 1975.

Palmer, S. E. Hierarchical structure in perceptual representation. Cognitive Psychology, 1977, 9, 441-474.

Palmer, S. E. What makes triangles point: Local and global effects in configurations of ambiguous triangles. Cognitive Psychology, 1980, 12, 285-305.

Palmer, S. E. Symmetry, transformation, and the strcture of perceptual systems. In J. Beck & F. Metelli (Eds.), Representation and organization in perception. Hillsdale, N.J.: Erlbaum, in press.

Palmer, S. E. Reference frames in shape perception. In preparation.

Palmer, S. E., & Bucher, N. M. Configural effects in perceived pointing of ambiguous traingles. Journal of Experimental Psychology: Human Perception & Performance, 1981, 7, 88-114.

Rock, I. Orientation and form. New York: Academic Press, 1973.

Shepard, R. N. & Judd, S. A. Perceptual illusion of rotation of three dimensional objects. Science, 1976, 191, 952-954.

Wallach H., & O'Connell, D.N. The kinetic depth effect. Journal of Experimental Psycholcgy, 1953, 45, 205-217.

Wertheimer, M. Untersuchungen zur Lehre von der Gestalt. Psychologishe Forschung, 1923, 4, 301-350. Translated in W. D. Ellis (ed.), A sourcebook of gestalt psychology. New York: Harcourt, Brace & Co., 1938.

Weyl, H. Symmetry. Princeton, N.J.: Princeton University Press, 1952.