Perceptual Simplicity and Modes of Structural Generation

Jacob Feldman

Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology,
Cambridge, MA 02139
jacob@psyche.mit.edu

Abstract

This paper describes a formal framework for perceptual categorization that can account for the salient qualitative predicates human observers are willing to ascribe to a closed class of objects, and consequently the simple groupings they can induce from small sets of examples. The framework hinges on the idea of a generative process that produces a given set of objects, expressed as a sequence of group-theoretic operations on a primitive element, thus ascribing algebraic structure to perceptual organization in a manner similar to Leyton (1984). Putatively, perceivers always seek to interpret any stimulus as a formally *generic* result of some sequence of operations; that is, they interpret each object as a typical product of some generative process.

The principle formal structure is a "mode lattice," which a) exhaustively lists the qualitative shape predicates for the class of shapes, and b) defines the inferential preference hierarchy among them. The mechanics are worked out in detail for the class of triangles, for which the predicted qualitative features include such familiar geometric categories as "scalene," "isosceles," and "right," as well as more "perceptual" ones like "tall" and "short." Within the theory it is possible as well to define "legal" vs. "illegal" category contrasts; a number of examples suggest that our perceptual interpretations tend to regularize the latter to the former.

Introduction

Among all the descriptions of a stimulus configuration that are in principle possible, human observers seem to favor the simplest ones, a generalization that can be traced back as least as far as the Gestaltists (e.g., Koffka, 1935; Köhler, 1947). Consequently, many perceptual theories have attempted to account for preferred descriptions by showing that they minimize some measure of complexity, often description length in some description language (e.g., Buffart, Leeuwenberg, & Restle, 1981; Darrell, Sclaroff, & Pentland, 1990; Hochberg & McAlister, 1953; Simon, 1972; see also Hatfield & Epstein, 1985; Leeuwenberg & Boselie, 1988; Perkins, 1976). This notion, often termed the "minimum principle," parallels the mathematical theory of informational complexity (e.g. Chaitin 1966, Kolmogorov, 1965), in which the complexity of a string is defined as the length of the shortest computer program that can generate it. However, these perceptual theories have often suffered from the fact that the description languages used must be somewhat arbitrary; the atoms of the language—individual features, predicates, and types of spatial relationships that are worth incorporating into descriptions—are difficult to motivate by underlying theory. Rather than finding one "minimum principle" underlying perception, as would be desired, perceptual theorists find many. Moreover, the relationship between descriptive simplicity and descriptive *correctness*—if there is any—has never been made completely clear.

This paper attempts to get around these problems by proposing a framework of stimulus description which explicitly generalizes across different ways of conceiving of the stimulus as having been constructed. A particular conception of the generating process underlying a particular class of stimuli entails a lattice relating the subspaces and half-spaces. This lattice both a) exhaustively enumerates the qualitative structural predicates and features of the class of objects, and b) makes the inferential preference relationships among them explicit. The scheme is worked out here in detail for one class of object, triangles; the resulting legal shape predicates seem, intuitively, both simple and natural. In this way the qualitative shape predicates that human observers are willing to ascribe to triangles, as evidenced in the categories they are able to induce from a handful of examples, are thus characterized formally, and (in theory) exhaustively.

Mode lattices

Leyton (1984) developed a view of perceptual interpretation in which a stimulus is treated by the observer as the product of a sequence of nested geometric transformations. Perceptual interpretation, then, proceeds by factoring the input, layer by layer, into successively simpler structures to which more complex ones are referenced. The algebraic structure this theory imposes on perceptual organization is intriguing and compelling—explaining, for example, why a parallelogram is perceived as a slanted rectangle, and a rectangle as a stretched square. A line segment, in this framework, might be the visible product of a translation operation on a point in the plane, which we label the origin (giving us translation invariance): $(\mathbf{0})\tau_{r,\theta}$. The space of all line segments constructed this way (the *configuration space*) has two dimensions, r and θ. A generic point in this space—a typical product of the generative process we are assuming to have produced all line segments—is a line with non-zero length that is neither exactly horizontal nor exactly vertical. Non-generic points fall on one of the subspaces (i.e., axes) of the space, and correspond to the class of vertical lines, the class of horizontal lines, and the origin point itself. Line segments that fall into one of these subspace classes are very special, and it is worthwhile making their specialness explicit in the perceiver's representation.

In general, we model each stimulus configuration u as the visible product of the action on the origin of a sequence of one-parameter operations

$$u = (0)\mathcal{T}_1^{\alpha_1} \cdot \mathcal{T}_2^{\alpha_2} \cdot \ldots \cdot \mathcal{T}_d^{\alpha_d},$$

where d, the number of one-parameter operations, is the dimensionality of the resulting configuration space, and each operation $\mathcal{T}_i$ is carried out to a magnitude α_i. Any operation whose magnitude is zero reduces to the identity operation, and thus drops out of the expression completely; this corresponds to a subspace of the configuration space. Within a given space (or subspace) the generic objects are all produced by generative strings in which none of the operations has dropped out. Recursively, a non-generic object that falls in a given subspace can either be generic in that subspace, or it can actually fall in a subspace of the subspace, in which case it is even more non-generic; that is, it has higher codimension in the overall configuration space. (The codimension is simply the difference between the dimension of an object and the dimension of the space in which it is embedded.) Each proper subspace divides each embedding superspace (of dimension one greater) in half; we call the two half-spaces *modes* of the space. Conversely, each (possibly improper) subspace of dimension k can be divided into two half-space modes by a subspace (of dimension k - 1) in k different ways.

Generic interpretations are preferred. Apparently, perceivers always seek to interpret the observed configuration as a *generic* product of some sequence of generating operations. The generating sequence of operations is pruned of identities until a generic sequence can be seen to yield the observed object; that is, if an observed object can be expressed with one of the operations as the identity, the object's interpretation drops down a dimensional level to an embedded subspace. Lower-dimensional (higher codimension) subspaces, since they have more dimensions fixed in a way unlikely to occur by accident, always make stronger inferences of structure. For a closed class of objects in one configuration space, therefore, all the predicates a perceiver is willing to apply to objects in the class should correspond to subspaces or half-spaces of the configuration space.

These spaces can all be connected up schematically in a "mode lattice." The mode lattice is built around the ordinary subspace lattice, to which is added, for each subspace, the two modal half-spaces of the embedding superspace into which the subspace divides it. The resulting construction makes the dimensionality and subspace relationships among interpretations of objects in the class explicit. Then, putatively, all allowable predicates on these objects can be read off the mode lattice in an automatic fashion. The next section describes a mode lattice for triangles.

Mode lattice for triangles

One way of conceiving a triangle is as two legs joined at a fulcrum, and a third leg joining their ends. We now cast this generating process formally as a sequence of operations on the origin, and construct the corresponding mode lattice. Note that any alternative way of formally conceiving the construction of triangles would lead to a different mode lattice, one which imposes a different geometry on the space of triangles. There are a finite number of such alternatives.

In this way of conceiving triangles, there are two controlling dimensions: the angle θ between the two sides that meet at the fulcrum, and the difference Δr between the lengths of the legs. Assuming scale invariance, the dimension of the configuration space d is thus 2, so that there are 3 (2^d - 1) proper subspaces: $r_1 = r_2$, $\theta = \pi/2$, and the origin, at which both $r_1 = r_2$ and $\theta = \pi/2$. ($\theta = \pi/2$ is an identity, briefly, because only this special angle differs from its complementary angle by the identity transform.) The mode lattice appears in Figure 1; the legend and explanatory schematics are on the left of the figure, and the lattice itself is drawn on the right. A dot has been drawn on all the sample triangles at the reference vertex, both in order to define it and to encourage the reader's perceptual apparatus to treat it as special, thus hopefully creating a bias towards seeing the triangles as products of a generative process of the sort described.

The "scalene" triangle (the term here meaning generic in the overall configuration space) appears at the top of the lattice. An arbitrary triangle, that is, a typical product of this generative process, is scalene. The proper subspaces are also familiar triangle categories: the axes are "isosceles" and "right", two categories traditionally treated specially, and the origin is of course "right isosceles." Necessarily, the origin triangle is shape-invariant, since there are no degrees of shape freedom left to vary. This triangle, via the spatial parameters it has fixed, in a sense implies the entire lattice.

The feature predicates that appear on the mode lattice should all be categories of triangles that human observers can induce easily from even a very small set of examples. This is most acutely seen when categories are contrasted with one another, as in a "Bongard" problem (Bongard, 1970; Feldman, in preparation; Richards, 1988). The next section proposes a definition of "legal" category contrasts, i.e., legal pairings of modes and subspaces from a mode lattice. The criterion for legal contrasts, though motivated on independent formal grounds, turns out to be intimately related to a simple model of naturally-occuring categories.

(a)

S 1 = generic ("scalene") triangles

S 2 = $\langle r_1 = r_2 \rangle$ = isosceles

S 3 = $\langle \theta = \pi/2 \rangle$ = right

S 4 = $\langle r_1 = r_2, \theta = \pi/2 \rangle$ = right isosceles

M 1 = M^{+}_{S2} = left leg longer

M 2 = M^{-}_{S2} = right leg longer

M 3 = M^{+}_{S3} = obtuse

M 4 = M^{-}_{S3} = acute

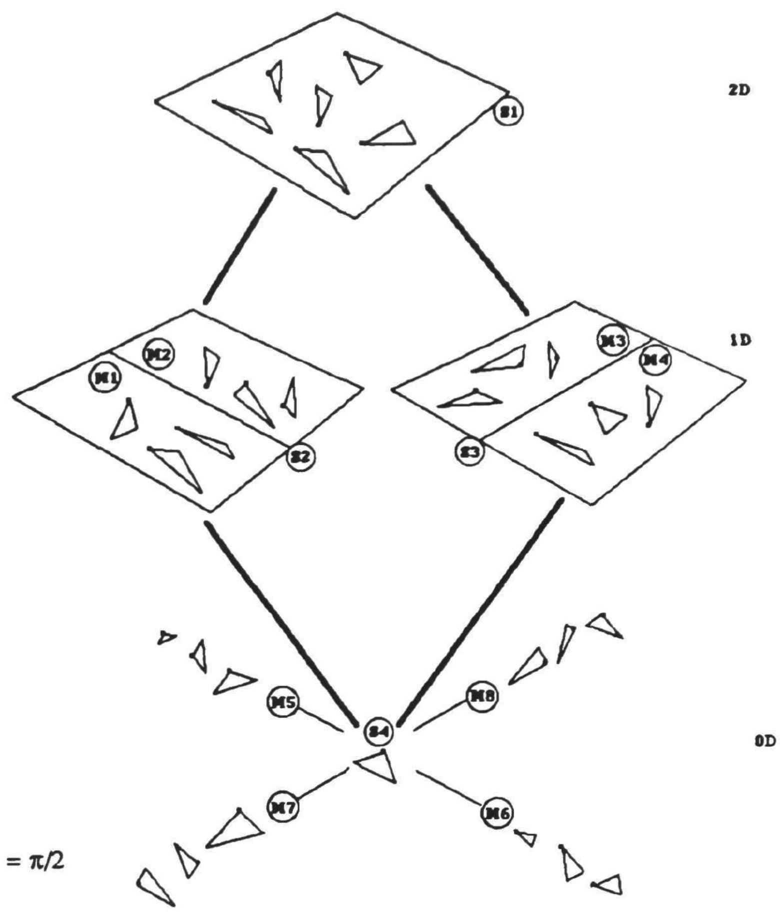
M 5 = M^{+}_{S4} (within S2) = obtuse isosceles

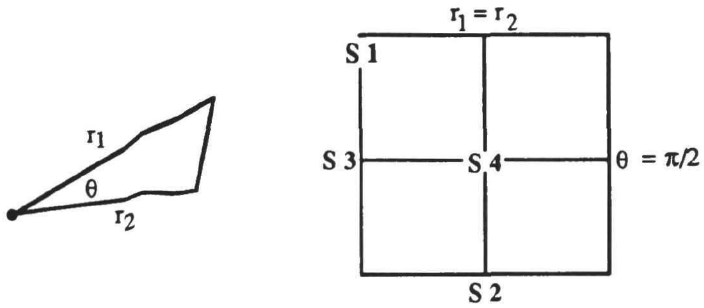
M 6 = M^{-}_{S4} (within S2) = acute isosceles

M 7 = M^{+}_{S4} (within S3) = right, with right leg longer

M 8 = M^{-}_{S4} (within S3) = right, with left leg longer

(b)





(c) (d)

Fig. 1. Mode lattice for triangles: (a) list of subspaces and modes (half-spaces); (b) mode lattice; (c) schematic of a triangle, labeling the angle and sides; and (d) schematic of the configuration space, showing the relation of the four subspaces. The notation M^{+}_{S2} (etc.) indicates the mode that is on the plus side of subspace S2.

Legal category contrasts

The 2-D triangle space has 4 subspaces (including the entire space) and 8 modes, so there are a total of 12 theoretically reasonable categories; in the absence of constraint any of the 66 pairings of these might make a good contrast. We now cut this number down by proposing constraints.

Legal category contrasts naturally fall into two types: between-space and within-space.

Between-space contrasts. Between-space contrasts, such as those between the two axes of the triangle space, or between the entire space and one axis, involve categories that can be written as completely different sequences of generative operations. A subspace of a space is generated by a subsequence of its defining operations, with the identity operation dropping out. Two crossing subspaces may be thought of as the product of two different operations on the same point. In either case, the resulting categories amount to different "species" of object; with different generative identities, a perceiver might justifiably expect them to have qualitatively different structural properties and attributes.

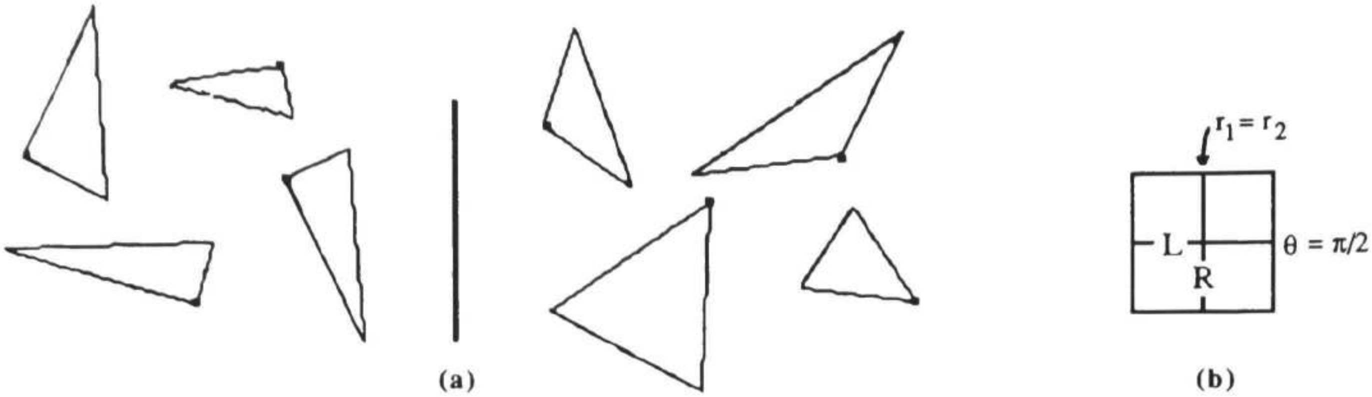


Fig. 2. (a) "Right" (L) vs. "Isosceles" (R) triangles; (b) schematic of the contrast, showing it as two subspaces of triangle configuration space.

An example appears in Figure 2. The reader may confirm that the category inductions are perceptually apparent and intuitively natural.

Within-space contrasts. A category within a (sub)space is a collection of quadrants of the space; any two such categories correspond to different classes within the same basic generating process. Though they cannot be expressed as the product of different sequences of operations, the contrasting categories manifest different values along some structural dimensions. That the differences between them are modal—they cross the boundaries delineated by the subspaces—suggests that such categories might, like between-space contrasts, exhibit qualitatively different structural properties or attributes.

To motivate the definition of legal within-space contrasts, consider that some collections of quadrants in a 2-D can be expressed as a contrast along one dimension that is then extruded symmetrically along the other dimension. For example, if the full 2-D space is divided in half, there is contrast between the two half-spaces across one axis (the bisecting one), but symmetry across the perpendicular axis. In such an arrangement, we call the asymmetrical axis the *contrast* component, and the symmetrical axis the *extrusion* component (because the contrast is extruded along it). Legal contrasts are defined to be those that can be expressed as a pure combination of an extrusion component of magnitude e and a contrast component of magnitude c, such that $c > 0$ (or else there is no contrast) and $c + e = d$ (all dimensions are accounted for). In the extreme case of pure contrast ($e = 0$), two diagonally opposite quadrants meet at the origin. In this case the contrast between the categories is complete: they differ on every essential structural parameter.

An example of a legal within-space contrast appears in Figure 3. Again, the contrast seems to capture a perceptually intuitive and salient distinction.

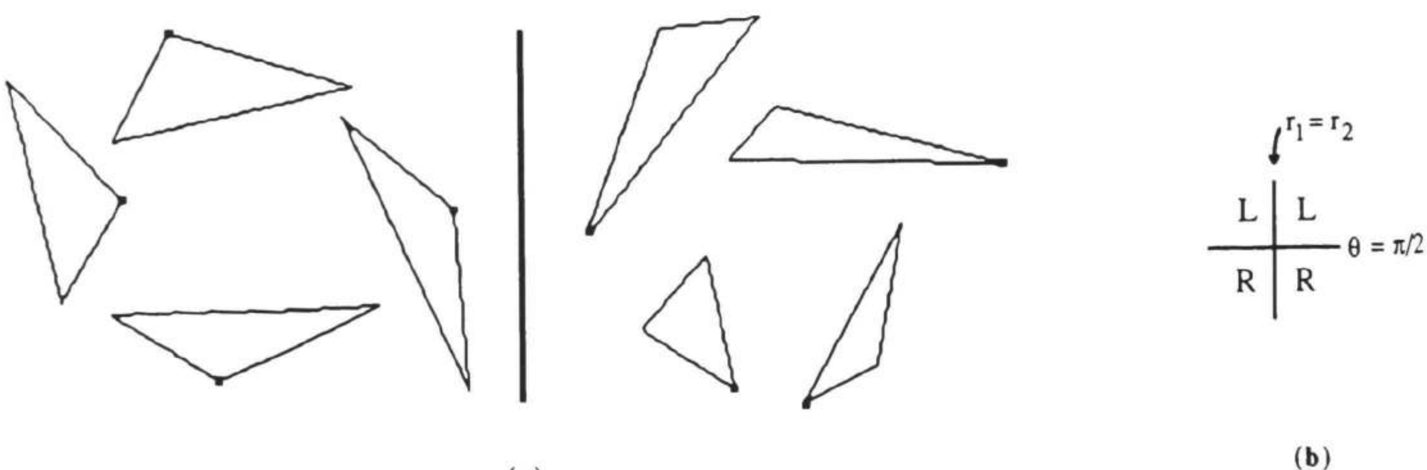


Fig. 3. (a) "Obtuse" (L) vs. "Acute" (R) main angle; (b) schematic of the contrast, showing it as two modal half-spaces of the configuration space. The categories are contrasted along the θ dimension, and extruded along the Δr dimension.

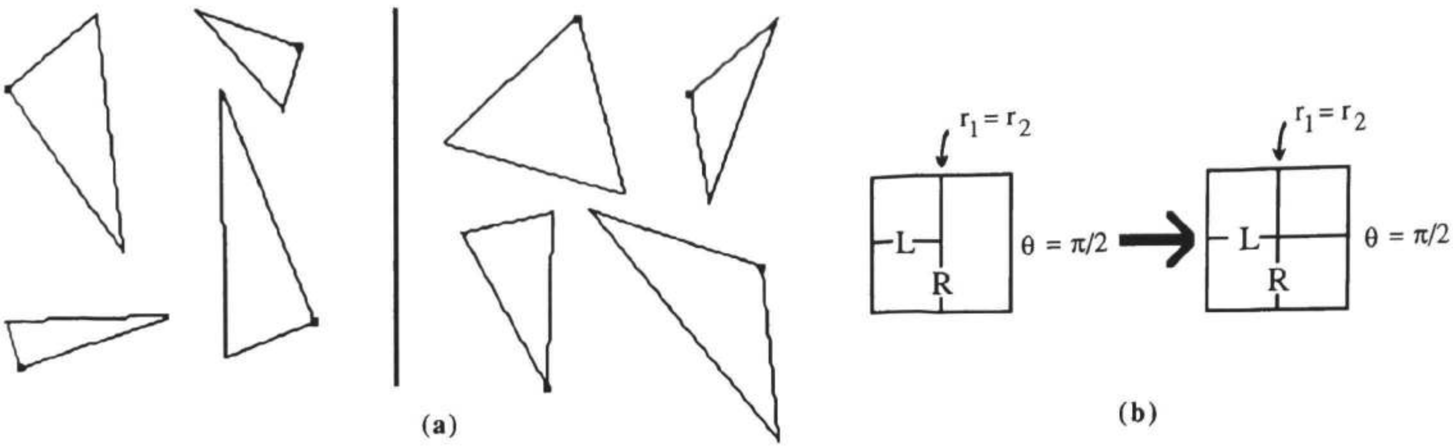


Fig. 4. (a) "Right & right leg longer" (L) vs. "Isosceles" (R); (b) schematic showing the regularization of the contrast to "Right" vs. "Isosceles."

These two examples are meant only to give the general flavor of legal contrasts: they capture in a formalism the intuitively natural contrasts that one may exhibit between triangles (at least those that were created by *this* generative process). There are other qualitative types of legal category contrast, for which space does not permit examples. All the basic types can be catalogued exhaustively, so that theory's categorization of hypothetically "natural" perceptual categories is complete.

Illegal category contrasts: regularization?

Non-modal contrasts. Many of the a priori possible category contrasts, both between- and within-space, are illegal as defined. Each of these, apparently, is regularized by our perceptual apparatus to one of the legal types. Figure 4 gives an example. The space of right triangles (L) is incomplete, including only triangles with the right leg longer, but our perceptual apparatus apparently either fails to notice this or considers it insignificant.

Amodal contrasts. So far, in considering possible classes of object within a space, we have completely ignored most possible subregions of the configuration space, in favor of the subspaces and modes. Nearly all subregions of the configuration space, of course, are not composed, even illegally, of subspaces and modes. Arbitrary contiguous patches of the space, not to even mention discontiguous parts (such as topologically disconnected or even nowhere continuous point sets), might in the absence of any constraint constitute reasonable stimulus categories. We would not expect our perceptual apparatus, however, to be able to correctly characterize world categories of arbitrary complexity. To reflect this, it should be the case that most topologically possible categories are *not* modal in any parameterization of the object class. Moreover, the total number of modal categories, considering all conceivable generative processes, should be finite. This turns out to be the case; all but a finite family of categories are amodal, that is, are not constructed from legal modal categories in any parameterization.

In pilot work we have looked at a particularly simple case of such amodal categories, which as such is a particularly good candidate to falsify the theory: a pair of simple linear categories that cross at a point (like the two axes of the configuration space), but whose crosspoint is not an origin in any parameterization (unlike the axes'). Such categories cross modal boundaries in a manner inconsistent with any legal category interpretation. The result is a pair of categories that seem not to have a perceptually reasonable distinction between them. When the categories are mixed together, moreover, the perceiver tends to regroup them modally. The "real" category is regularized to one which is incorrect but is at least perceptually reasonable.

Conclusion: legal categories as world categories

Though the term "qualitatively similar" has some intuitive meaning to human perceivers, a formal definition is elusive; formally, *any* two distinct items in a set share the same number of properties unless some theory constrains the properties (Watanabe, 1985). Formal theories of perception (Bennett, Hoffman, & Prakash, 1989; Bobick & Richards, 1986; Feldman, 1991; Jepson & Richards, 1991; Leyton, 1984, 1986, 1988, 1989; see also Fell, 1976; Witkin & Tenenbaum, 1983, 1986) attempt, in essence, to distinguish the perceptual predicates that perceptual systems such as ours favor from the much wider class of arbitrary predicates, nearly all of which are perceptually useless. The mode lattice theory provides a formal definition, which the examples suggest matches human intuitions.

We may now reasonably ask, why does this definition work for the human perceptual system? In particular, do "legal" categories correspond reliably to extant natural categories in the physical and biological world? It turns out that they do. It can be shown that an extremely simple, formal model of natural categories and processes reduces exactly to the legal categories, as defined above. The model, briefly, assumes that natural categories manifest heavy internal structure within themselves, and correspondingly sharp contrasts (that is, sharper than a priori necessary) between each other. This idea is a particularization of the Principle of Natural Modes, articulated by Bobick (1987) and Richards & Bobick (1988). Space requires that a full discussion, including the requisite formal definition of "structure," be left to a future paper. The thrust is that the apparent psychological preference for modal categories, as evidenced by the various types of regularization demonstrated above, seems to be rooted in sensible inferential logic.

Perceptual interpretations, it has long been assumed in perceptual theories, should (1) be as simple as possible consistent with the image data, and (2) accurately represent the essential structural features of the object. The theory outlined in this paper sheds light on what these two desiderata have to do with each other, and suggests why achieving the second might require achieving the first. If the perceiver is able to identify a formal model of the generation of the stimulus class that more or less correctly mirrors the actual physical process by which the stimulus object was created, then the lowest-dimension description possible within that generation scheme, read directly off the mode lattice—in a very particular sense, the simplest description possible—will accurately represent the structurally important properties of the object.

Acknowledgements

I wish to thank Don Hoffman, Jan Koenderink, Shimon Ullman and most especially Whitman Richards for much helpful criticism of the ideas expressed in this paper.

References

Bennett, B., Hoffman, D. & Prakash, C. 1989. *Observer mechanics*. London: Academic Press.

Bobick, A. 1987. Natural object categorization. Technical Report no. 1001, M.I.T. A.I. Lab.

Bobick, A. & Richards, W. 1986. Classifying objects from visual information. A.I. Memo no. 879, M.I.T. A.I. Lab.

Bongard, M. 1970. *Pattern recognition*. New York: Spartan Books.

Buffart, H., Leeuwenberg, E., & Restle, F. 1981. Coding theory of visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 7 (2):241-270.

Chaitin, G. 1966. On the length of programs for computing finite binary sequences. *Journal of the Association for Computing Machinery*, 13 (4):547-569.

Darrell, T. Sclaroff, S. & Pentland, A. 1990. Segmentation by minimal description. Proceedings of the Third International Conference on Computer Vision, Osaka, Japan. Los Alamitos, CA: IEEE Computer Society Press.

Feldman, J. 1991. Perceptual decomposition as inference: Continuous curvilinear processes. Occasional Paper no. 42, Center for Cognitive Science, M.I.T.

Feldman, J. in preparation. Perceptual induction in Bongard problems.

Fell, H. 1976. Detectable properties of planar figures. *Information and Control*, 31 (2):107-128.

Hatfield, G. & Epstein, W. 1985. The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin*, 97(2):155-186.

Hochberg, J. & McAlister, E. 1953. A quantitative approach to figural "goodness." *Journal of Experimental Psychology*, 46:361-364.

Jepson, A. & Richards, W. 1990. What is a percept? Occasional Paper no. 43, Center for Cognitive Science, M.I.T.

Koffka, K. 1935. *Principles of Gestalt psychology*. New York: Harcourt, Brace and Co.

Köhler, W. 1947. *Gestalt psychology*. New York: Liveright Publishing Co.

Kolmogorov, A. 1965. Three approaches to the quantitative definition of information. *Problems of Information Transmission* 1(1):1-7.

Leeuwenberg, E., & Boselie, F. 1988. Against the likelihood principle in visual form perception. *Psychological Review*, 95:485-491.

Leyton, M. 1984. Perceptual organization as nested control. *Biological Cybernetics*, 51:141-153.

Leyton, M. 1986. A theory of information structure. I. General principles. *Journal of Mathematical Psychology*, 30:102-160.

Leyton, M. 1988. A process-grammar for shape. *Artificial Intelligence*, 34:213-247.

Leyton, M. 1989. Inferring causal history from shape. *Cognitive Science*, 13:357-387.

Perkins, D. 1976. How good a bet is good form? *Perception*, 5:393-406.

Richards, W. 1988. *Natural computation*. Cambridge: M.I.T. Press.

Richards, W. & Bobick, A. 1988. Playing twenty questions with nature. In Z. Pylyshyn ed. 1988, *Computational processes in human vision: an interdisciplinary perspective*. Norwood, NJ: Ablex.

Simon, H. 1972. Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 79(5):369-382.

Watanabe, S. 1985. *Pattern recognition: Human and mechanical*. New York: John Wiley & Sons.

Witkin, A. & Tenenbaum, J. 1983. On the role of structure in vision. In H. Rosenfeld ed. 1983, *Human and machine vision*. New York: Academic Press.

Witkin, A. & Tenenbaum, J. 1986. On perceptual organization. In A.P. Pentland ed. 1986, *From pixels to predicates: Recent advances in computational and robotic vision*. Norwood, N.J.: Ablex.